CENTRE
DE RECHERCHE
SUR LES TRANSPORTS
(CRT)
CENTRE
FOR RESEARCH
ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE

MÉMOIRE D'ÉTUDIANT

C.P. 6128, SUCCURSALE A
MONTRÉAL (QUÉBEC)
CANADA
H3C 3J7

TÉLÉPHONE : (514) 343-7575
TÉLEX : 05-24146 BIBPOLYTEC
TÉLÉCOPIEUR (FAX) : (514) 343-7121

CRT-650

LES COÛTS DU TRANSPORT FERROVIAIRE PAR CONTENEURS EN FRANCE : ÉTUDE DE CAS*

par

Christel Vasseur

Rapport de stage fin d'études d'ingénieur de l'École Centrale Paris 1987-1988, co-dirigé par les Professeurs Pierre Dejax, Département des Sciences Economiques, Sociales et Humaines, Ecole Centrale de Paris, et Teodor G. Crainic, Département de Sciences administratives, Université du Québec à Montréal. Le stage a eu lieu au Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, dans le cadre du projet de coopération France-Québec No. 20020686 FQ.

Université de Montréal-HEC-POLY

Centre de recherche sur les transports - Publication #650

Octobre 1989

90-01403/

Dépôt légal - 4e trimestre 1989

REMERCIEMENTS

Avant de commencer ce rapport je tiens à remercier toutes les personnes qui m'ont permis d'effectuer ce stage lié à la Recherche Opérationnelle: Mr Pierre DEJAX de l'Ecole Centrale de Paris, Mrs Teodor CRAINIC et Louis DELORME de l'Université de Montréal, ainsi que toutes les personnes travaillant au CRT.

Je remercie également la Commission Franco-Québécoise, pour sa contribution financière à cette étude (projet de collaboration n°200-206-86-FQ).

I.INTRODUCTION 2

II.LE PROBLEME DE LA COMPAGNIE DE TRANSPORT ET L'ETUDE EFFECTUEE EN 1987 3
A.Le problème de la compagnie 3
B.L'étude effectuée en 1987
1.Objectifs de l'étude 4
2.Modélisation 4
3.Résultats obtenus 5

III.PRESENTATION DE CETTE ETUDE 7
A.Objectifs 7
B.Méthodologie adoptée 8

IV.REALISATION DE LA NOUVELLE BASE DE DONNEES 9
A.Nécessité d'un traitement des données 9
B.Hypothèses retenues 10
C.Fusion distancier-fichier d'observations. 10

V.REGRESSIONS 11
A.Identification de la fonction 11
1.Variables explicatives et groupes. 11
2.Groupe d'étude 12
3.Différentes régressions testées. 13
a.Fonction linéaire simple
b.Utilisation de ratios
c.Transformation de la variable expliquée DEP
d.Transformation des variables expliquée et explicatives
e.Transformation du groupe étudié
4.Comparaison pour RNEP / CH / FCL entre les deux modèles logarithmiques 23
5.Découpages en groupes: contraintes à respecter 26
6.Utilisation de la taille-type pour la partition en groupes 27
7.Remarques 29
B.Généralisation aux autres groupes 30
1.Groupes retenus 30
2.Résultats 31
a.Coefficients
b.Principales statistiques
c.Intervalles de confiance de l'estimation pour chaque groupe
d.Etude des résidus et résidus

VI.CONCLUSION 39

BIBLIOGRAPHIE 41

ANNEXES

I.INTRODUCTION

La planification du transport de conteneurs pour le commerce international est une activité très complexe étant donné le contexte général de compétitivité, la qualité de service à rendre au client, le nombre de données à prendre en compte,... Mais l'enjeu économique est immense pour les entreprises concernées.

A partir de données fournies par une grande compagnie de transport maritime pour l'année 1986, le LEIS (Laboratoire Economique, Informatique et Social) de l'Ecole Centrale de Paris et le CRT (Centre de Recherche sur les Transport) rattaché à l'Université de Montréal ont proposé un plan tactique qui réduisait les coûts de 41 % (soit plusieurs dizaines de milliers de francs) par rapport aux coûts observés.

Ce chiffre, même s'il doit être considéré comme une économie "idéale" (l'économie effective sera probablement inférieure vu toutes les dépenses imprévisibles qui surgiront lors du passage à la réalité opérationnelle) suffit à nous convaincre de l'utilité économique de la planification.

Cependant une étude plus poussée est encore nécessaire, pour rendre le programme d'optimisation utilisé plus proche de la réalité opérationnelle, mais également pour augmenter ses performances. Un axe de recherche particulier est l'obtention d'un "graphe complet", où tous les arcs (liaisons) entre les différents noeuds (points) du réseau de transport de l'entreprise soient affectés d'un coût de transport et donc utilisables par le programme. En effet , seuls les arcs utilisés en 1986 le sont pour le moment.

L'objet de ce rapport est de déterminer ce coût dans le cas particulier du transport par rail, à l'intérieur de la France et mettre au point une méthodologie applicable dans les autres cas.

II. LE PROBLEME DE LA COMPAGNIE DE TRANSPORT ET L'ETUDE EFFECTUEE EN 1987

A. Le problème de la compagnie

Cette compagnie fournit à ses clients des conteneurs de différents types où sont chargées les marchandises à transporter.

Elle doit également amener les conteneurs pleins jusqu'à ses clients, les transporter une fois remplis, puis les renvoyer à un port ou dépôt après livraison. Le nombre et le type de conteneurs pour chaque client est relativement bien connu, de même que la destination.

En 1986 la plupart des conteneurs une fois arrivés à destination, livrés aux clients et vidés, étaient renvoyés, au port d'origine, quelque soit sa distance. Ceci entraînait de nombreux transports à vide, inutiles et surtout très coûteux.

B.L'étude effectuée en 1987

1.Objectifs de l'étude

Cette étude (P.Dejax, T.Crainic, L.Delorme, et coll.,1987) était destinée à montrer l'utilité pour la compagnie de transport d'adopter un "plan tactique", à moyen terme, et de se doter des outils correspondants: base de données et logiciels d'optimisation des flux de conteneurs.

Limitée dans un premier temps à la planification des flux de conteneurs vides en fonction des mouvements de transports de conteneurs chargés, l'étude a permis:

1/ d'une part de déterminer des schémas de flux (à moyen terme) à adopter entre les clients, les ports et les dépôts intérieurs qui aideront ensuite à prendre les bonnes décisions au niveau opérationnel en ce qui concerne l'affectation des conteneurs vides aux différents dépôts,

2/ d'autre part, d'évaluer différents "scénarios du futur". Ces scénarios peuvent par exemple porter sur la répartition des modes de transport, la politique commerciale, l'évolution de la demande ou des coûts maritimes, ou la coopération avec les compagnies maritimes confrères.

2.Modélisation

La modélisation nécessite trois étapes préalables:

1/ Génération du réseau, ses "sommets" (i.e. les clients offreurs et demandeurs de conteneurs, et les ports et dépôts, puis ses arcs (i.e. les liaisons permises).

2/ Calcul des coûts, sur chaque arc permis, soit en reprenant le coût observé, soit en le calculant.

3/ Détermination des besoins et des disponibilités de chaque client, et dépôt.

Ensuite il est possible d'effectuer la planification proprement dite à l'aide d'algorithmes de recherche opérationnelle adaptés (T.G.Crainic, P.Dejax, L.Delorme, 1986) qui a donné des résultats tout-à-fait encourageants.

3.Résultats obtenus

Les tableaux suivants montrent la part consacrée à chaque type de mouvement, et le pourcentage (maximum) économisé lorsque l'on a recours à la planification tactique.

Mouvements observés en 1986:

Mouvement		Part du coût total
Clients	Dépôts/Ports	22 %
Dépôts/Ports	Dépôts/Ports	39 %
Dépôts/Ports	Clients	36 %
Clients	Clients	3%

(P.Dejax, T.Crainic, L.Delorme, et coll.,1987)

Mouvements optimisés:

Mouvement		Part du coût total
Clients	Dépôts/Ports	42 %
Dépôts/Ports	Dépôts/Ports	7 %
Dépôts/Ports	Clients	51 %

(P.Dejax, T.Crainic, L.Delorme, et coll.,1987)

Comparaison des coûts globaux entre l'optimisation et les mouvements observés en 1986:

	Réel 86	Optimisation
Plan Tactique	ø	59
Réalité Opérationnelle	100	?

Ce tableau montre la variation de coûts qui résulte d'une bonne planification des transports de conteneurs vides: étant données les demandes en 86, pour 100 F effectivement dépensés (en ne tenant compte que de l'opérationnel) l'utilisation d'un plan tactique prévoit une dépense limitée à 59 F

Cependant on ne peut considérer l'économie de 41 % réalisée comme "certaine" (d'où le point d'interrogation):

* le passage du niveau tactique (objectif à atteindre) au niveau opérationnel (réalité quotidienne) s'accompagne automatiquement d'une hausse des coûts, dûe aux dépenses imprévisibles.
L'économie réelle serait donc inférieure à 41 %

* le programme de planification utilisé n'est pas parfait et peut encore être amélioré ce qui conduira, selon les modifications apportées, soit à des économies plus importantes soit à des prévisions plus exactes (plus proche de l'opérationnel).

III. PRESENTATION DE CETTE ETUDE

A. Objectifs

Nous avons vu que l'une des étapes de la modélisation consistait à calculer les coûts de tous les arcs "permis". Or pour l'étude faite en 1987 nous ne disposions que de "coûts observés", relatifs à des arcs déjà utilisés en 1986 ce qui nous a obligés à considérer tous les autres arcs comme "interdits".

Cette démarche est acceptable dans une première approche, car elle fait l'hypothèse, assez réaliste, que la compagnie utilisait les arcs les plus intéressants, et que les arcs non utilisés étaient peu rentables. Cependant une planification rigoureuse ne peut s'en contenter: il est possible que certains arcs soient laissés de côté simplement parce que "l'habitude veut que..."

Cette étude se propose donc de voir dans quelle mesure on peut affecter un coût à des arcs inexplorés:

* en les calculant numériquement dans le cas du transport par rail à l'intérieur de la France (i.e. en déterminant une fonction f qui donne pour chaque mouvement son coût, connaissant ses différentes caractéristiques)

* en proposant une méthodologie applicable aux autres modes (routes, barge,...), aux autres pays, ...

B.Méthodologie adoptée

Le principe de l'étude est d'utiliser le fichier 1986 des mouvements de conteneurs de la compagnie pour déterminer, à l'aide de modèles de régression, si le coût observé pour un mouvement quelconque peut se déduire simplement de ses caractéristiques (qui restent à identifier) par une "fonction de coût" puis utiliser la fonction obtenue afin de calculer les coûts inconnus (mouvements non utilisés en 86).

Cette "fonction de coût", si elle existe, sera probablement de la forme:

Coût = f (Distance, Poid, ...)

où f varie avec la taille des conteneurs transportés,
leur type (isotherme, découvert,...),
le mode de transport (rail, route,...),
le mode de facturation (transport à la charge du transporteur ou du client,...),
le nombre de clients desservis,
etc..

Il nous faut donc identifier trois éléments distincts mais étroitement liés les uns aux autres:

* les groupes homogènes au sein desquels on a une unique fonction f
* les variables explicatives
* la forme des fonctions f (linéaire, ...) pour chaque groupe.

Pour nous affranchir de l'interdépendance groupe/variables/forme nous avons successivement

1/ determiné les variables susceptibles d'expliquer la dépense, a priori, afin de constituer un fichier de travail sur lequel effectuer les régressions

2/ identifié la forme de la fonction et les variables explicatives en essayant puis testant plusieurs régressions pour un groupe homogène particulier (obtenu en fixant toutes les variables discrètes),

3/ partitionné les observations en groupes en respectant les contraintes d'homogénéité et d'effectif suffisant décrits plus loin (paragraphe V.A.S.), en faisant l'hypothèse que la fonction est la même pour tous les groupes, aux coefficients près,

4/ analysé la fonction de coût obtenue dans chaque groupe pour déterminer quelle confiance accorder aux résultats (en particulier étude des résidus).

IV. REALISATION DE LA NOUVELLE BASE DE DONNEES

Cette base de données doit contenir pour chaque observation toutes les variables susceptibles d'expliquer la dépense.

Nous avons retenu un certain nombre de variables mentionnées dans le fichier 86 des mouvements observés par la compagnie (les variables rejetées sont les libellés clients ou dépôts, la suite des opérations effectuees pour un même client, le type de mouvement: import ou export, commercial ou technique). Toutes ces variables peuvent être considérées comme "variables de décisions".

Nous y avons ajouté une variable exogène: la distance, issue d'un distancier rail SNCF. (C'est d'ailleurs l'utilisation de ce distancier qui nous oblige à limiter notre étude aux mouvements par rail à l'intérieur de la France)

A. Nécessité d'un traitement des données

Nous disposions des données observées en 1986 par la compagnie de transport mais les distances mentionnées ne pouvaient pas être utilisées dans la recherche de la fonction de coût pour deux raisons:

1/ Ces distances n'étaient pas toujours fiables, certains transports avec même origine et destination étaient affectés d'une distance de plusieurs centaines de kilomètres.

2/ L'objectif de l'étude est avant tout de trouver le coût des arcs non utilisés en 86, aucun enregistrement de la compagnie ne mentionne donc leur distance. Pour la connaître il faudra se referrer à un fichier externe.

Aussi avons-nous fait appel à la SNCF. Mais le distancier obtenu utilisait une codification différentes: la première tâche à effectuer était de transformer les libellés SNCF en numéros de cantons administratifs, comme dans le fichier de la compagnie.

B.Hypothèses retenues

Le distancier SNCF comportait 394 gares identifiées par leur nom. Après les avoir repérées sur une carte administrative et avoir noté le numéro du canton auquel elles appartenaient on a obtenu

* 4 villes étrangères (Bâle, Kehl, Genève, Vintimile)
* 26 gares "introuvables", soit trop petites pour être mentionnées sur la carte, soit dotées d'un libellé SNCF qui ne correspond pas au nom de la commune administrative à laquelle elle appartient.

De plus certains cantons possédaient plusieurs gares mentionnées dans le distancier. Suivant les cas on a retenu

* la distance à la gare principale, quand son importance est telle qu'on peut négliger les autres,
* ou la moyenne arithmétique des distances à chaque gare du canton (par exemple pour toutes les gares entourant Paris) ce qui conduit à des distances non nulles à l'intérieur d'un même canton.

Nous obtenons ainsi un distancier-rail entre cantons.

C.Fusion distancier-fichier d'observations de la compagnie

La fusion des deux fichiers, le fichier de la compagnie (observations 86) et le distancier-rail se fait simplement en ajoutant à chaque enregistrement, mode RNEP (Rail National Embranchement Particulier) ou RNRR (Rail National, combiné Rail-Route), une donnée supplémentaire, KMDIS, qui est la distance lue dans le distancier, entre le canton de départ et le canton d'arrivée.

La description du fichier final, OBSFRD, figure en annexe page 13.

V.REGRESSIONS

A.Identification de la forme de la fonction

L'objectif était d'obtenir une fonction f donnant le coût d'un arc à partir d'un certain nombre de données "pertinentes". Il faut rappeler que la fonction f n'est pas forcément la fonction f réelle (d'ailleurs existe-t-elle?) mais une fonction suffisamment proche des données pour pouvoir être utilisée dans l'algorithme de planification tactique. En effet notre approche n'est pas comptable mais stratégique.

1.Variables explicatives et groupes

L'examen des données contenues dans OBSFRD amène à chercher f telle que l'on ait

DEP = f(POID, CONT, KMDIS, EXPE)

où CONT est le nombre de conteneurs transportés
EXPE le nombre de clients desservis,

et f varie suivant TATI (taille et type de conteneur),
REGD (région de départ),
REGA (région d'arrivée),
MODE (rail national uniquement ou
combinaison rail et route),
RE (Carrier Haulage ou Formule Mixte),
F (plein, vide, ou semi-plein).

(cf. annexes B et C pour plus de détails)

Enfin la variable retenue pour représenter la distance est KMDIS, issue du distancier, et non la distance issue du fichier de la compagnie pour plusieurs raisons:

Nous avons déjà parlé de l'objectif de l'étude (à savoir le coût des arcs inutilisés), et des erreurs de saisie du fichier de la compagnie.

De plus la proximité des deux distances (en annexe des graphes montrent leur relativement bonne concordance) et des essais de régression faits suivant chaque point de vue (distancier ou observations de la compagnie) montrent que les deux sont équivalents (voire KMDIS meilleure)

2.Groupe d'étude

Pour trouver la forme de f nous avons essayé plusieurs régressions sur une sous-population a priori homogène, puisque toutes les variables discrètes ont été fixées:

TATI = 1	20 pieds 9m3
REGD = 1	Paris / Nord Pas de Calais
REGA = 1	
MODE = 3	RNCP
RE = 1	Carrier Haulage
F = 1	FCL (plein)

Ce groupe comprend 270 observations, ce qui est suffisant pour donner un sens statistiques aux regressions testées.

Le logiciel utilisé pour la régression est BMDP (W.J.Dixon, University of California, 1981). La selection de la fonction se fait en deux étapes:

* Un premier "tri grossier" permet de rejeter rapidement plusieurs régressions, les deux principaux critères de selection étant

1/ le R^2 (coefficient de corrélation, significatif de l'utilisation faite des données, un R^2 proche de 1 montre une régression utilisant pleinement toutes les données).

2/ le graphe des résidus versus la dépense estimée (ils doivent, pour que la régression soit valable former une "bande" horizontale, preuve de leur caractère aléatoire).

Il faut remarquer que dans cette phase "d'identification" nous pouvons nous permettre de régresser suivant plusieurs variables, même si certaines sont redondantes ou inutiles. Dans le pire des cas l'équation obtenue leur affectera un coefficient négligeable.

* Une étude plus approfondie des fonctions retenues, où nous examinons les résultats des tests de normalité des résidus, de nullité de leur moyenne, etc... décrits en annexes.

Lors de cette seconde phase il est nécessaire de bien identifier quelles sont les variables explicatives car dupliquer une variable(en utilisant deux variables très corrélées) ou ajouter une variable inutile réduit considérablement l'intervalle de confiance de la variable expliquée.

3.Différentes régressions testées

Le logiciel de régression utilisé (BMDP) ne permet, pour des études multidimensionnelles (plus d'une variable explicative Xi), que des modèles linéaires.

Aussi avons-nous essayé des fonctions linéaires (du type Y = Σ bi Xi) ou transformables en fonctions linéaires (l'utilisation de logarithmes permet d'identifier les coefficients d'une fonction de type produit).

Plusieurs fonctions ont ainsi été testées:

a.Fonction linéaire simple

DEP = a + b POID + c CONT + d KMDIS

Une régression pas à pas montre que toutes les variables explicatives sont nécessaires.
R^2 = 85 %
mais le graphe des résidus vs DEP n'est pas acceptable, car trop "évasé" (cf. figure 1).

b.Utilisations de ratios

DEP/KMDIS = a + b POID + c CONT + d KMDIS R^2= 6 %

DEP/CONT = a + b POID + c CONT + d KMDIS R^2= 35 %

DEP/CONT = a + b POID/CONT + c CONT + d KMDIS R^2= 37 %

DEP/(POID*KMDIS) = a + b POID + c CONT + d KMDIS
+ e CONT/POID + f CONT/KMDIS

Cette régression "pas à pas" fait rentrer a, d, e, f, et donne un R^2 de 41 % et les variables sont très corrélées.

Diviser la dépense par la distance correspond à la tarification appliquée par certaines compagnies de transport (paiement d'un fixe + prix du Km). Il en est de même pour le poid (compagnies de déménagement), ou le nombre de conteneurs (chaque conteneur serait payé indépendamment des autres). Ce raisonnement amène à considérer une multitude de combinaisons.

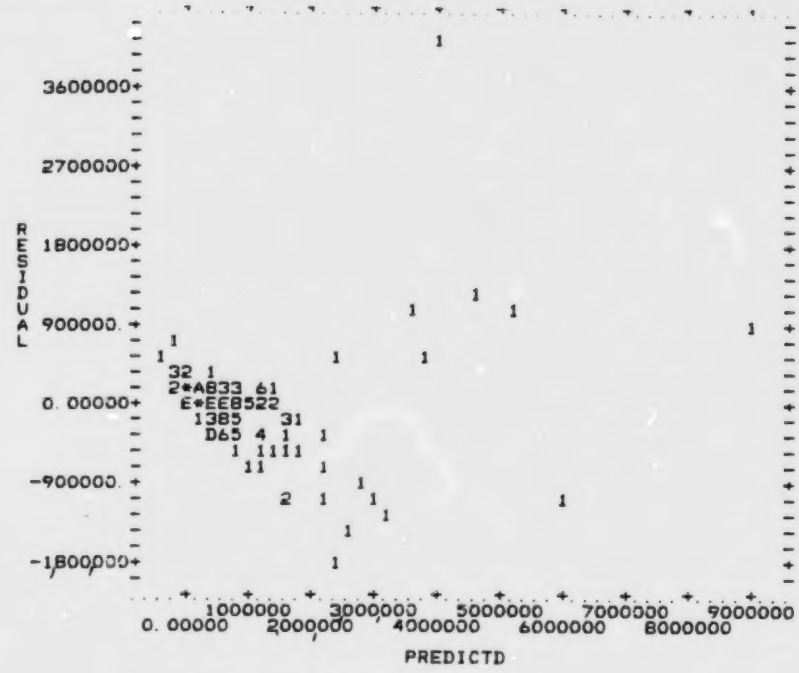


Figure 1

RNEP CH FCL TATI = 1 REGA=1

DEP = a + b CONT + c POID + dKMDIS

PAGE 7 BMDP1R REG 8

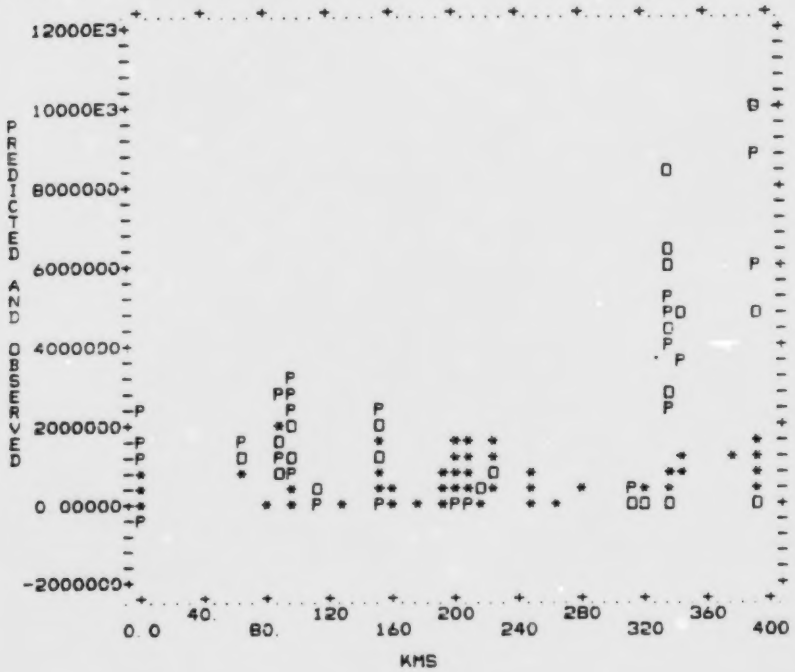


Figure 2

RNEP CH FCL TATI = 1 REGA = 1

DEP = a + b CONT + c POID + d KMDIS

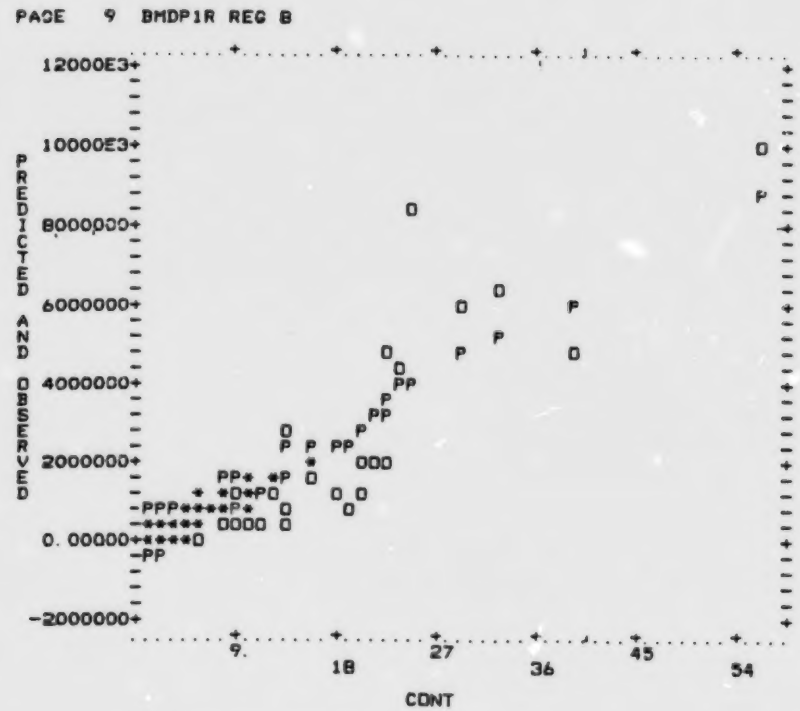


Figure 3

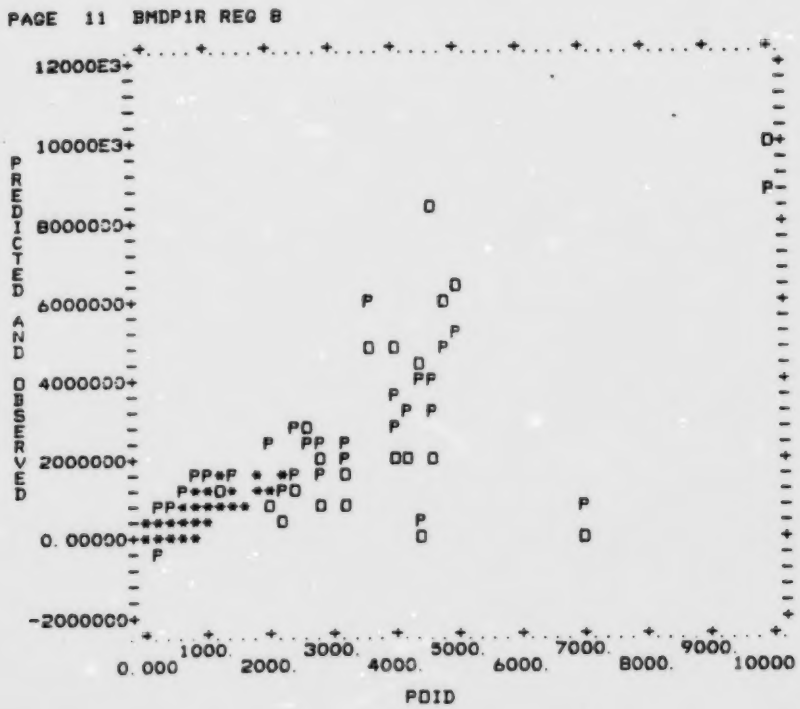
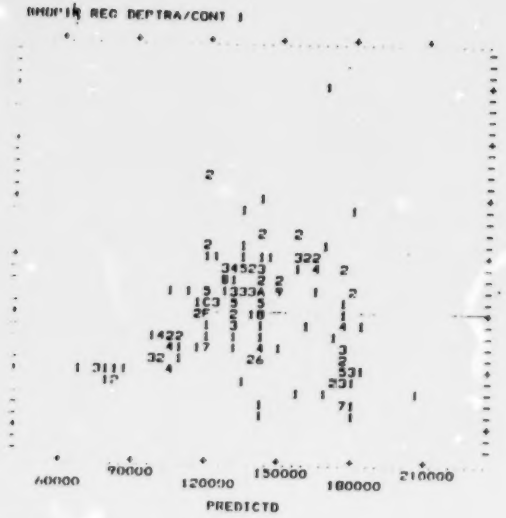
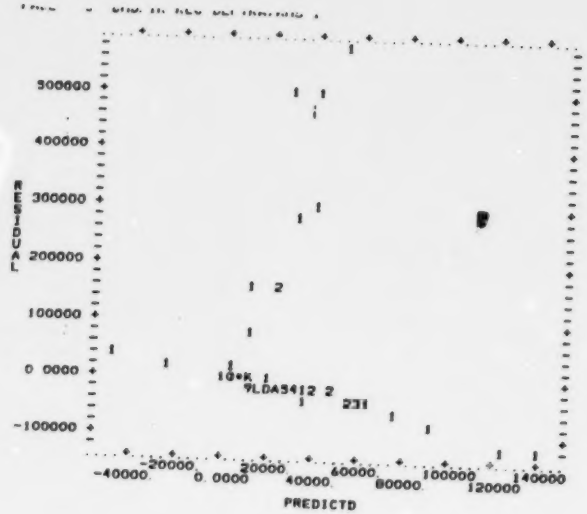


Figure 4



DEP/CONT = a + b KMDIS + c CONT + d POID

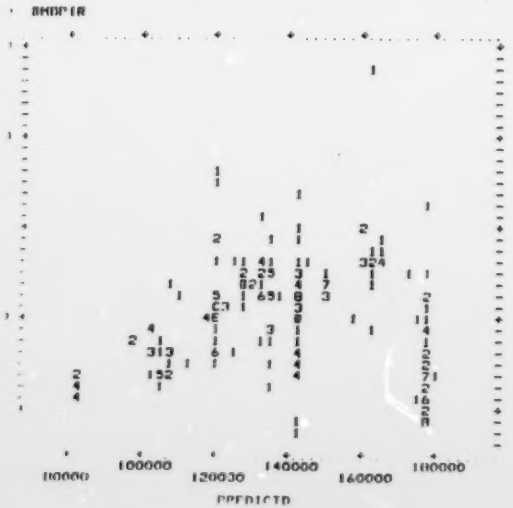


DEP/KMDIS = a + b KMDIS + c CONT + d POID

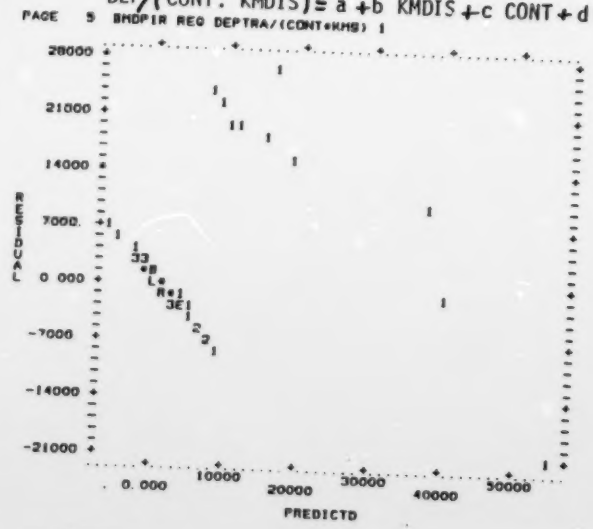
Figures 5 et 6

RNEP FCL CH TATI = 1 REGA = 1

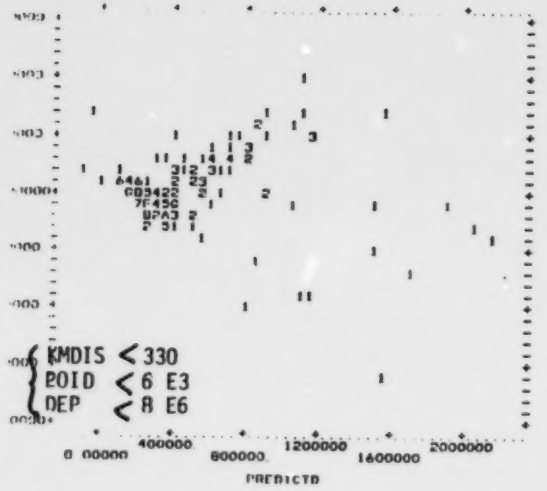
DEP/CONT = a + b KMDIS + c CONT + d POID/CONT



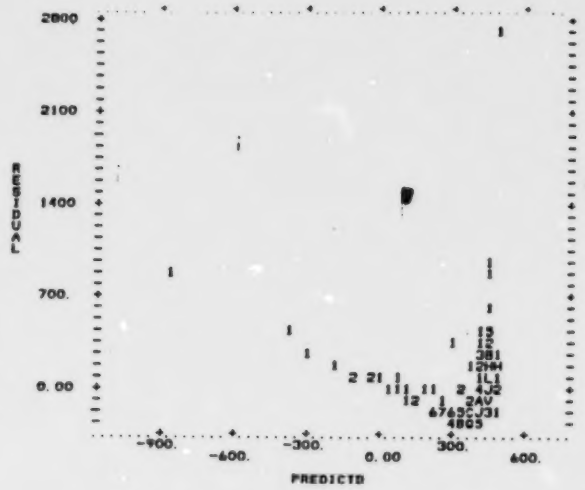
DEP/(CONT. KMDIS) = a + b KMDIS + c CONT + d POID + e POID/CONT + f POID/KMDIS



Figures 7 et 8



DEP = a + b KMDIS + c CONT + d POID

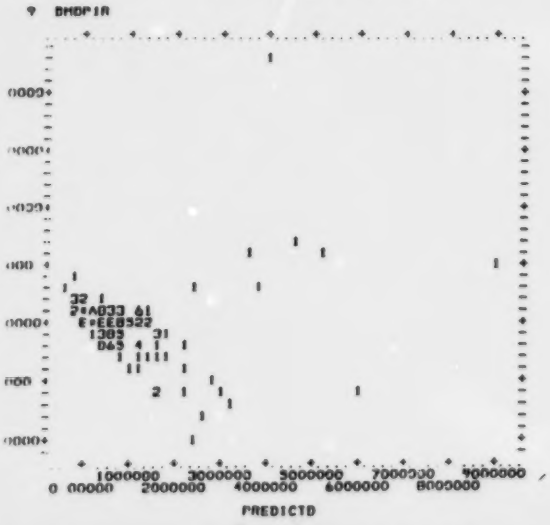


1/ DEP = a + b KMDIS + c CONT + d POID

Figures 9 et 10

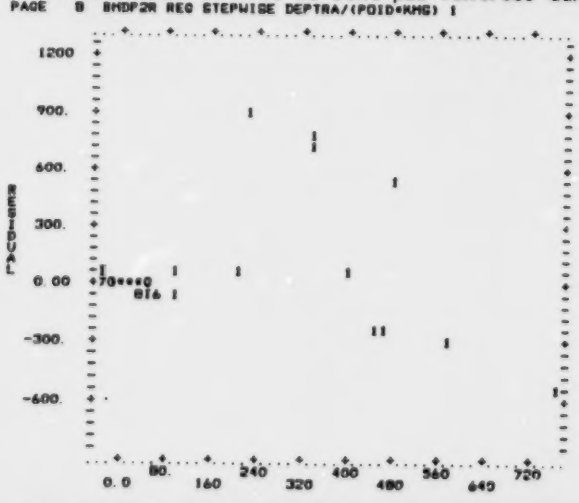
CH FCL RNEP TATI = 1 REGA = 1

DEP = a + b KMDIS + c CONT + d POID



DEP/(KMDIS . POID) = a + b KMDIS + c CONT/POID + d CONT/KMDIS

(Les var. CONT et POID ne sont pas rentrées dans le rég. pas a pas)



Figures 11 et 12

Ainsi jusqu'ici aucune n'est réellement satisfaisante. En effet, même si le R^2 est parfois bon (85 %) les résidus sont alors "évasés" (cf. figures 5, 6, 7, 8, et 12): ils augmentent quand les dépenses augmentent ce qui est caractéristique d'une régression pour laquelle il faudrait

* soit pondérer les enregistrements (mais comment? le nombre des observations nous empêche de les pondérer par la variance : cela obligerait à inverser une matrice carrée de dimension 2000 * 2000 pour certains groupes puis de la multiplier avec une autre...),

* soit supposer une relation non linéaire mais nous retrouvons le problème du logiciel utilisé...(cf.V.A.3),

* soit transformer les variables.

C'est donc cette dernière méthode que nous avons utilisée pour la suite.

c.Transformation de la variable expliquée DEP

DEP' = a + b POID + c CONT + d KMDIS R^2= 63 %

DEP' = ln(DEP)	R^2= 63 %	
DEP' = √DEP	R^2= 79 %	
DEP' = √DEP	R^2= 75 %	(cf. figures 10, 13, 15)
DEP' = 1/DEP	R^2= 23 %	
DEP' = DEP ^{2}	R^2= 63 %	

Nous pouvons également introduire la constante apparue lors de la régression linéaire (a), qui est assimilable à un coût fixe:

DEP' = ln(DEP + Constante)	R^2= 78 %	
DEP' = 1/(DEP + Constante)	R^2= 56 %	(cf. figures 14, 16)

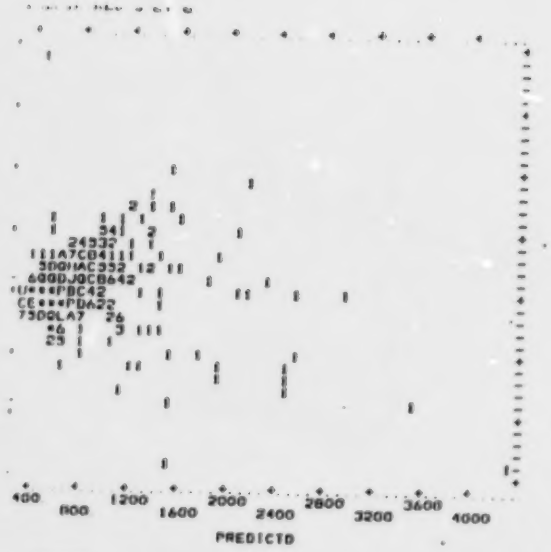
Mais les graphes (résidus vs DEP) ne sont toujours pas acceptables.

d.Transformation de la variable KMDIS

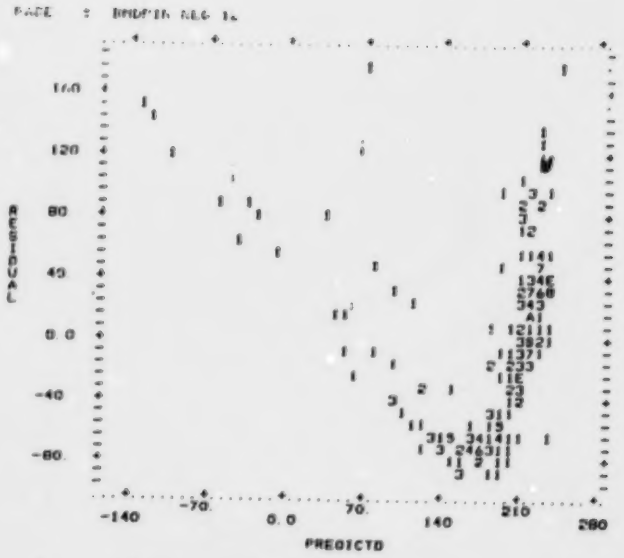
DEP = a + b POID + c CONT + d KMDIS'

avec KMDIS'= √ KMDIS	R^2= 84 %	
ln(KMDIS)	R^2= 85 %	
ln(KMDIS)^2	R^2= 86 %	

Mais là encore les graphes des résidus ne sont jamais satisfaisants (cf. figures 17 à 20).



$\sqrt{DEP} = a + b\ KMDIS + c\ CONT + d\ POID$

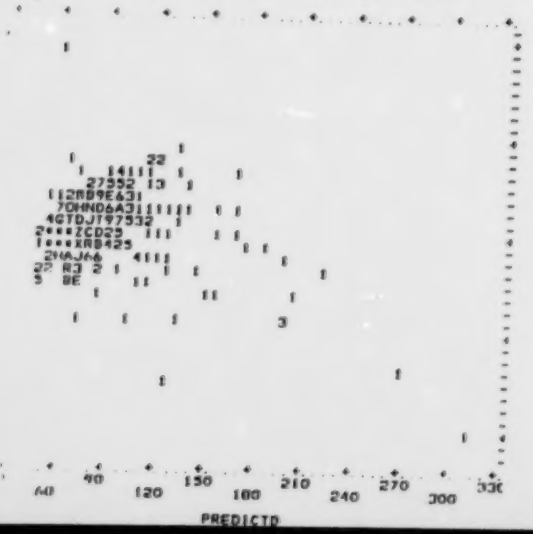


$1/(DEP + Cte) = a + b\ KMDIS + c\ CONT + d\ POID$

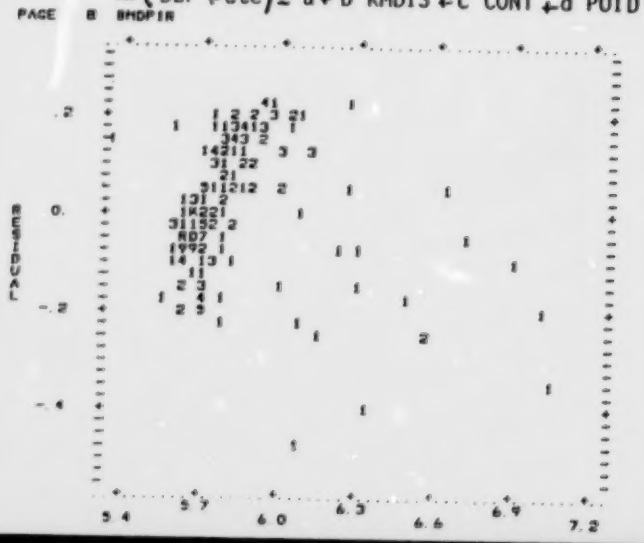
Figures 13 et 14

RNEP FCL CH TATI = 1 REGA = 1

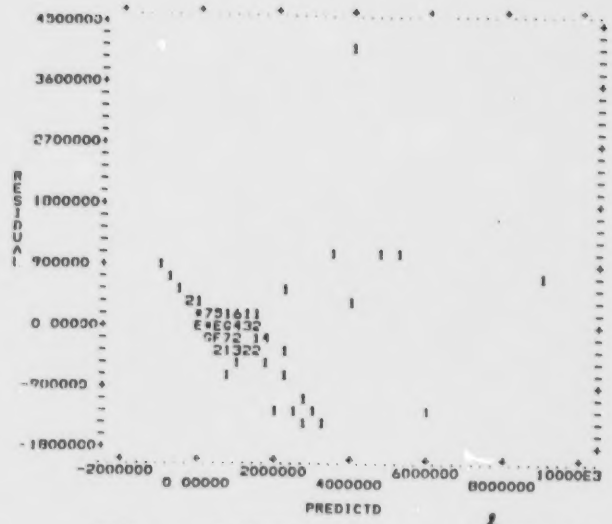
$\sqrt[3]{DEP} = a + b\ KMDIS + c\ CONT + d\ POID$



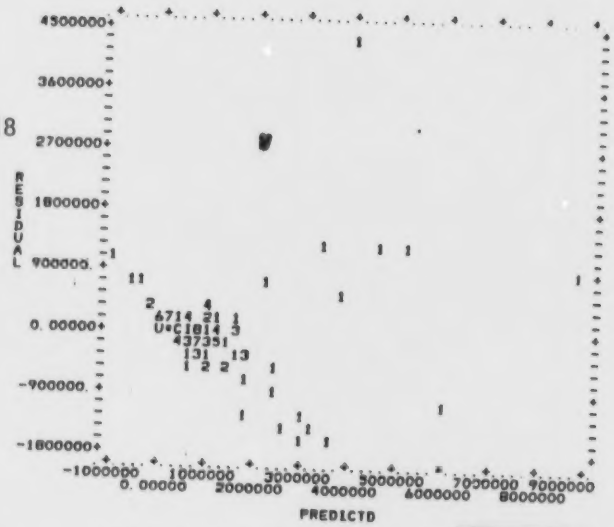
$\ln(DEP + Cte) = a + b\ KMDIS + c\ CONT + d\ POID$



Figures 15 et 16



Figures 17 et 18



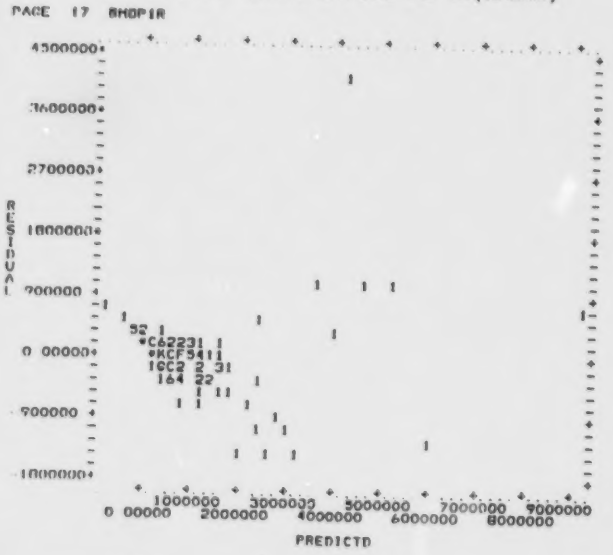
DEP = a + b CONT + c POID + d $\ln(KMDIS)^2$

DEP = a + b CONT + c POID + d $\sqrt{\ln(KMDIS)}$

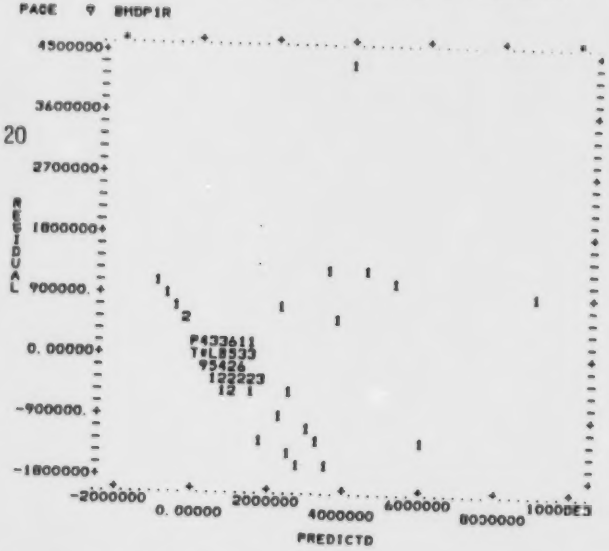
RNEP FCL CH TATI = 1 REGA = 1

DEP = a + b CONT + c POID + d $\ln(KMDIS)^3$

DEP = a + b CONT + c POID + d ln(KMDIS)



Figures 19 et 20



e.Transformation des variables expliquée et explicatives

$\ln(DEP) = a + b \ln(POID) + c \ln(CONT) + d \ln(KMDIS)$ R^2=90 %

Le graphe des résidus (présenté en annexe pour les groupes définitifs) est bien en forme de bande. Cette régression peut donc être étudiée plus en détail.

f.Transformation du groupe étudié

Enfin nous avons étudié si le groupe considéré était réellement homogène.

L'examen des graphes des résidus vs variables (cf figures 2, 3, et 4) a permis de fixer des bornes au-delà desquelles les observations semblaient peu utilisables (car elles correspondaient probablement à des transports exceptionnels dont le coût est négocié au "coup par coup" avec le transporteur), ceci en terme de poids maximal, dépense maximale, ou kilométrage maximal . Mais la régression obtenue après élimination des valeurs "limites" n'est pas concluante (cf. figure 9).

Un dernier type de régression est obtenu en séparant les enregistrements suivant des tranches de Km:
de 0 à 50 Km
de 50 à 100 Km
de 100 à 150 Km etc...

Cet essai était motivé par les systèmes de tarification couramment utilisé par les sociétés de transport (location de véhicules, déménagement, ..) mais également par l'allure particulière du graphe DEP vs KMS.

Dans ce cas la simple régression linéaire sans transformation est améliorée, bien que les résidus soient encore légèrement évasés. Mais c'est encore la régression où toutes les variables, expliquées et explicatives, sont ln-transformées qui donne les meilleurs résultats.

Enfin, nous avons réessayé quelques fonctions décrites en 1/ 2/ et 3/ pour des tranches kilométriques qui se révélaient homogènes dans les cas linéaires et logarithmiques (cf tranche 300-350 Kmdis) sans que cela se revèle intéressant (les R^2 sont toujours aussi faible ce qui montre que la mauvaise adaptation de ces fonctions provient de f et non pas d'une mauvais regroupement des données).

Donc après cette première "étape éliminatoire" on peut retenir deux régressions :

$$\ln(DEP) = a + b \ln(POID/CONT) + c \ln(CONT) + d \ln(KMDIS)$$

et $$\ln(DEP) = a' + b' \ln(POID/CONT) + c' \ln(CONT)$$

où a', b', c', varient selon la tranche de km dans laquelle on se situe.

L'utilisation du log-népérien est un moyen de linéariser un produit et donc d'utiliser le logiciel de régression pour déterminer les coefficients A, b, c, et d.

$$DEP = A\ (POID/CONT)^{b}\ (CONT)^{c}\ (KMDIS)^{d}$$

où $A = \exp(a)$

4.Comparaison pour RNEP / CH / FCL entre les deux modèles avec ou sans tranche de Km

(1) ln(DEP) = a + b ln(POID/CONT) + c ln(CONT) + d ln(KMDIS)

(2) ln(DEP) = a' + b' ln(POID/CONT) + c' ln(CONT)

où a', b', c', varient selon la tranche de Km dans laquelle on se situe.

Deux éléments distincts doivent être pris en compte pour déterminer les tranches de Km optimales: la taille du groupe ainsi défini (il faut suffisamment d'observations pour obtenir une régression fiable) et l'homogénéité de ce groupe.

Le programme BMDP1R permet de vérifier ce dernier point. En effet il est capable d'effectuer une régression, d'abord sur toutes les observations, puis en distinguant différents groupes prédéfinis, pour finalement procéder au test F définissant l'intérêt de séparer tous ces groupes (cf annexe).

Une première régression a été faite en distinguant des tranches de 50 ou 100 Km:

Nous trouvons, pour l'équation (2), les coefficients suivants:

Intervalle	a' cste	b' poico	c' cont	R^2
0 - 50 Km	10.957	.117	.627	.718
50 - 100 Km	10.622	.156	.945	.971
100 - 150 Km	11.053	.144	.751	.794
150 - 200 Km	12.336	-.102	.951	.963
200 - 250 Km	11.310	.122	.895	.928
250 - 300 Km	10.091	.385	.786	.785
300 - 350 Km	10.438	.339	.889	.829
350 - 400 Km	11.062	.215	.922	.916
400 - 500 Km	11.299	.223	.737	.789
500 - 600 Km	10.354	.442	.878	.933
600 - 700 Km	11.408	.243	.854	.899
700 - 800 Km	11.779	.169	.809	.838
800 - 900 Km	11.177	.295	.967	.932
900 - 1100 Km	11.522	.253	.877	.896
≥ 1100 Km	12.165	.154	.924	.962

L'intervalle de confiance à 95 % pour les coefficients de la régression étant typiquement de 2 fois leur écart-type, les valeurs de ce tableau peuvent être modulées de

$\pm$ .141 pour b'
$\pm$ .114 pour c'

Ces deux valeurs représentent la somme des écart-type maximum et minimum des différents groupes:

$$2\,[(\,\sigma\ \text{min} + \sigma\ \text{max}\,)\ /\ 2]$$

Ensuite des tests (BMDP1R) ont été fait pour regrouper les catégories qui ne présentaient pas de grandes différences. En effet un trop grand nombre de tranches distinctes nuit à la qualité de la régression (le nombre d'observation devient insuffisant) et à la facilité d'utilisation des résultats.

Nous obtenons les ratios suivants pour l'utilité de la séparation des groupes:

	F-ratio	Prob
100 - 150 - 200	17.461	.00000
250 - 300 - 350	18.806	.00000
300 - 350 - 400	2.286	.07876
600 - 700 - 800	3.102	.02618

Un F-ratio élevé est significatif de groupes trop différents pour pouvoir être fusionnés.
Prob est la probabilité (calculée à partir du F-ratio) pour que les groupes soient identiques.

Ceci montre que l'on ne peut pas réduire le nombre de groupes et que l'on est obligés d'en garder 14 pour la seule catégorie RNEP / FCL / CH / 20 Pieds!

C'est pourquoi il est préférable de s'en tenir au modèle (1) qui sera nettement plus facilement exploitable.

D'ailleurs distinguer de très nombreuses tranches revient finalement à rentrer KMDIS dans l'équation, l'intérêt étant de pouvoir estimer la dépense en ne connaissant la distance que de façon approximative.

5.Découpage en groupes

Il doit satisfaire à deux conditions essentielles:

a/ Homogénéité des observations au sein d'un même groupe ce qui nous amène a priori à distinguer

le régime	CH / FM
le remplissage	FCL / LCL / VIDE
le mode	RNEP / RNRR

b/ Effectif suffisant (N ≥ 50) pour que toutes les régressions aient un sens ce qui oblige

d'une part à ne pas retenir toutes les tailles-types (12 au total) mais à simplement distinguer les 20 pieds des 40 pieds,

d'autre part à prendre toutes les régions confondues (hypothèse justifiée par le fonctionnement au niveau national de la SNCF). Le tableau en apporte la confirmation. Il donne le nombre des mouvements RNEP / CH / FCL / 20' effectués entre les différentes régions (décrites en annexe page 14).

Mouvements RNEP FCL CH 20'

D\A	1	2	3	4	5	6	7	TOT
1	624	2	16	192	280	153	123	1390
2	45			1	3			49
3	49			2	10			61
4	58			78	171		5	312
5	22	1	12	77	51	12	45	220
6	95			9	223			327
7	260			116				376
TOT	1153	3	28	359	854	165	173	2735

6.Utilisation de la taille-type pour la partition en groupes

Les deux tableaux suivants nous montrent la répartition des types et tailles de conteneurs pour chaque type de mouvement. Ils nous montrent que pour respecter le critère d'"effectif suffisant" dans le groupement des observations on ne peut considérer chaque taille-type séparément.

TAILLE - TYPE	1	2	3	4	5	6	20'
RNEP / CH / FCL	1659	9	16	31	3	1040	2758
LCL	167			1			168
VIDE	1938	16	33	69	44	1107	3207
FM / VIDE	113		3	3		502	621
RNRR / CH / FCL	1052	1	13	3		503	1572
VIDE	2		1				3

TAILLE - TYPE	7	8	9	10	11	12	40'
RNEP / CH / FCL	15	76	1	240	456	1	789
LCL					40		40
VIDE	42	103	2	306	646	20	1119
FM / VIDE	1			348	43		392
RNRR / CH / FCL	16	1		111	443	4	575

Comment les fusionner?

La signification des six catégories 20 pieds (en particulier 1- 20'usage général, 6- divers 20') nous a amenés à toutes les confondre. De même pour tous les 40 pieds.

La question suivante est de savoir s'il faut également confondre 20' et 40'. BMDP effectue un test pour savoir si deux groupes doivent être séparés. Une régression a ainsi été faite sur les deux groupes,

RNEP / CH / FCL / 20 pieds
RNEP / CH / FCL / 40 pieds

Les résultats indiquent:

2 groupes réunis R^2 =.847

ln(DEP) = 8.110
+ .852 ln(CONT)
+ .269 ln(POID/CONT)
+ .489 ln(KMDIS)

	Intervalle à 95%
20 pieds	
ln(DEP) = 7.508	[7.361,7.656]
+ .862 ln(CONT)	[.848 , .875]
+ .263 ln(POID/CONT)	[.244 , .282]
+ .578 ln(KMDIS)	[.561 , .595]
40 pieds	
ln(DEP) = 7.527 ln(CONT)	[7.313,7.741]
+ .943 ln(CONT)	[.924 , .962]
+ .187 ln(POID/CONT)	[.161 , .213]
+ .673 ln(KMDIS)	[.644 , .702]

Il faut donc effectivement étudier séparément les deux tailles de conteneurs.

7.Remarques

* Nous faisons l'hypothèse (faute d'observations) qu'il est inutile de distinguer les régions et les tailles-types. C'est la qualité des résultats obtenus qui nous permettra de savoir si elle est justifiée.

* Toutefois on ne peut se passer de distinguer les transports 20 pieds ou 40 pieds (1 ≤ TATI ≤ 6 ou 7 ≤ TATI ≤ 12), les régimes (Carrier Haulage ou Formule Mixte), les vides des pleins, le rail simple RNEP et le rail-route RNRR.

En effet le test effectué par BMDP pour savoir si deux groupes doivent être séparés ou s'ils peuvent être regroupés montre que

RNEP	CH / FM	VIDE	20'	ne peuvent être regroupés
RNEP	CH	FCL / LCL	20'	———
RNEP / RNRR	CH	FCL	20'	———

* Afin d'avoir des variables explicatives peu corrélées entre elles il est préférable de travailler sur POID/CONT, qui représente le poid moyen du conteneur, variable indépendante du nombre de conteneurs.

* Travailler sur le logarithme au lieu des variables brutes, i.e. sur un produit au lieu d'une somme, est satisfaisant pour la signification de l'équation car "intuitivement" on s'attend à ce que les différentes variables ne se juxtaposent pas mais aient des effets combinés. Cependant ceci conduit à une exploitation plus difficile des résultats car l'intervalle de confiance sera calculé sur ln(DEP) et non sur DEP or

$$\ln(A) - \alpha \leq \ln(DEP) \leq \ln(A) + \alpha$$

ssi

$$1 / \exp(\alpha) \leq DEP / A \leq \exp(\alpha)$$

ssi

$$A / \mu \leq DEP \leq A * \mu \qquad \text{où } \mu = \exp(\alpha)$$

B.Généralisation des résultats aux autres groupes:

1.Groupes retenus

Finalement nous retenons les populations suivantes:

MODE	REGIME	PLEIN/VIDE	20 Pieds	40 Pieds	TOTAL
RNEP	CH	FCL	2758	789	3547
		LCL	168	40	208
		VIDE	3207	1119	4326
	FM	VIDE	621	392	1013
RNRR	CH	FCL	1572	575	2147
		VIDE	3	0	0
TOTAL			8329	2915	11244

Ces chiffres correspondent aux populations "initiales", ils ne tiennent pas compte des observations erronées. Dans la régression finale certaines observations, manifestement erronées, ont été écartées.

En effet BMDP signale trois observations particulières:
la plus éloignée du nuage de points etudiés (distance de Mahalanobis maximale)
la plus mal expliquée par la régression (résidu maximum)
la plus influente (distance de Cook maximale)

Nous avons ainsi pu vérifié dans le fichier OBSFRD leur vraisemblance, et rejeter certaines.

On obtient finalement des effectifs quelque peu inférieurs après avoir éliminé:

* pour tous les groupes KMDIS et KMCGM < 10

* RNEP / CH / FCL / 20 POID = 7040
  / 40 DEP = 3636600 et 25800
  VIDE / 20 KMDIS < 30
  / 40 KMDIS < 30
  RNRR / CH / FCL / 20 DEP = 496900
  / 40 DEP = 5918000

Mais il est évident que certaines observations aberrantes subsistent: faute d'être vraiment très éloignées des autres ou de jouer un rôle très importants dans la détermination des coefficients elles n'ont pas pu être détectées (et d'ailleurs était-ce nécessaire, vu leur peu d'influence sur le modèle ?).

Enfin le programme d'analyse des résidus produit un diagramme où le résidu pour chaque cas est porté pourvu qu'il soit inférieur en valeur absolue à deux écarts-types. Il donne également le pourcentage de ces "données hors-limites" parmi les observations traitées.

2.Résultats:

Les régressions effectuées pour les 5 groupes distingués dans le paragraphe précédent utilisent l'équation

$$\ln(DEP) = a + b \ln(CONT) + c \ln(POID/CONT) + d \ln(KMDIS)$$

(Les 4 groupes "vide" correspondent à des enregistrements dont le poid est nul donc n'admettant pas de coefficient c)

a.Coefficients

	a	b	c	d
RNEP / CH / FCL / 20	7.5085	.8618	.2632	.5780
/ 40	7.5269	.9430	.1872	.6733
LCL / 20	8.5503	.8844	.0516	.5196
/ 40	5.5211	.9826	.0915	1.0696
VIDE / 20	10.0363	.7990		.3084
/ 40	10.1051	.9472		.3506
FM / VIDE / 20	10.0827	.8681		.2885
/ 40	10.3543	.9883		.3343
RNRR / CH / FCL / 20	10.4982	.8899	.1842	.2247
/ 40	10.1738	.9691	.0709	.3941

La comparaison de ces résultats amène un certain nombre de remarques sommaires:

- la variation de la constante a quand on passe de RNEP(rail) à RNRR(rail-route) rend probablement compte des coûts de camionnage et chargement/déchargement

- l'influence du nombre de conteneurs semble être à peu près la même pour tous les groupes (même ordre de grandeur de b)

- en revanche le poid intervient beaucoup moins pour LCL (remplissage partiel) que FCL (plein). Peut-être serait-il plus pertinent dans le cas FCL de considérer le volume ?

- lors de transport à vide, la distance parcourue est moins influente pour la Formule Mixte (FM) que le Carrier Haulage (CH). Ceci nous insiterait donc à préférer, lors des longs trajets, la FM au CH.

b. Principales statistiques:

		R^2	$\sigma(\tau)$	$\sigma(\tau)/y$	out.	N
RNEP / CH / FCL	20	.883	.2707	2.07%	3.40%	2735
	40	.942	.1938	1.46%	2.54%	786
LCL	20	.900	.2752	2.11%	5.36%	168
	40	.980	.1207	0.94%	5.00%	40
VIDE	20	.874	.3031	2.38%	4.16%	2859
	40	.933	.2443	1.89%	3.51%	968
FM / VIDE	20	.794	.2340	1.87%	6.20%	500
	40	.884	.1740	1.37%	3.55%	366
RNRR / CH / FCL	20	.846	.3086	2.32%	3.41%	1552
	40	.790	.2183	1.64%	4.45%	562

R^2 représente le taux d'utilisation des données.

$\sigma(\tau)$ est l'écart-type de la variables estimées [$\tau \approx \ln(DEP)$]
$\sigma(\tau)/y$ est une première approximation du pourcentage d'erreur dans l'estimation (écart-type/moyenne)

out. est le pourcentage de résidus supérieurs à 2 écarts-types en valeur absolue.

N est le nombre final d'observations du groupe

(cf.Annexe A pour plus de détails)

c.Intervalles de confiance de l'estimation pour chaque groupe:

Les régressions faites fournissent une estimation τ de ln(DEP). En annexe A.5.d on voit que l'on peut obtenir un intervalle de confiance pour ln(DEP) par la formule

$$\tau \pm 2 [s \sqrt{X'CX} + \sigma(a)]$$

où s est le carré moyen des résidus.
X vecteur des moyennes des observations
$\sigma(a)$ écart-type de la constante

Le tableau suivant donne le résultat de ce calcul pour chaque groupe

	Intervalle de confiance pour DEP
RNEP / CH / FCL / 20	.853 exp(τ) < DEP < 1.173 exp(τ)
/ 40	.799 exp(τ) < DEP < 1.252 exp(τ)
LCL / 20	.363 exp(τ) < DEP < 2.756 exp(τ)
/ 40	.523 exp(τ) < DEP < 1.913 exp(τ)
VIDE / 20	.909 exp(τ) < DEP < 1.100 exp(τ)
/ 40	.853 exp(τ) < DEP < 1.172 exp(τ)
FM / VIDE / 20	.818 exp(τ) < DEP < 1.222 exp(τ)
/ 40	.839 exp(τ) < DEP < 1.192 exp(τ)
RNRR / CH / FCL / 20	.707 exp(τ) < DEP < 1.414 exp(τ)
/ 40	.658 exp(τ) < DEP < 1.520 exp(τ)

d.Etude des résidus et résidus orthogonalisés:

* Test de la moyenne nulle des résidus orthogonalisés (cf annexe A.3.b.) :

Pour chaque groupe la statistique t permet de donner un pourcentage de confiance pour le calcul de
[X-bar / (s / √N)]

RNEP / CH / FCL / 20 t(75%) = .674 ≤ t = .734 < t(90%) = 1.282
RNEP / CH / FCL / 40 t = .220 < t(75%) = .674
RNEP / CH / LCL / 20 t(75%) = .674 < t = .779 < t(90%) = 1.282
RNEP / CH / LCL / 40 t(90%) = 1.303 < t = 1.340 < t(95%) = 1.684)
RNEP / CH / VIDE / 20 t = .303 < t(75%) = .674
RNEP / CH / VIDE / 40 t = .244 < t(75%) = .674
RNEP / FM / VIDE / 20 t = .408 < t(75%) = .674
RNEP / FM / VIDE / 40 t(75%) = .674 < t = .963 < t(90%) = 1.282
RNRR / CH / FCL / 20 t(90%) = 1.282 < t = 1.345 < t(95%) = 1.645
RNRR / CH / FCL / 40 t(99.5%) = 2.576 < t = 2.885 < t(99.75%) = 3.291

Ainsi le meilleur niveau auquel on puisse accepter l'hypothèse d'une moyenne nulle est

Groupe	Niveau Max.	Moy.Calculée
RNEP / CH / FCL / 20	75 %	-.00380
/ 40	REJET	-.00153
LCL / 20	75 %	-.01676
/ 40	90 %	.02667
VIDE / 20	REJET	-.00217
/ 40	REJET	.00396
FM / VIDE / 20	REJET	-.00428
/ 40	75 %	.00880
RNRR / CH / FCL / 20	90 %	-.01055
/ 40	99.5 %	-.02649

Le rejet de l'hypothèse d'une moyenne nulle malgré une valeur faible de la moyenne calculée provient de la multiplication, dans le test, par $\sqrt{N}$.

* Tests de normalité

Le graphe "Detrended Normal Probability Plot of Residuals" indique pour chaque groupe l'écart qui existe entre les résidus obtenus par la régression et la loi normale la plus proche, abstraction faite de l'effet de tendance (on a ramené la pente à zéro). Ils montrent un défaut de normalité que les tests numériques confirment.

La description des statistiques de Kolmogorov-Smirnov, de Cramer-von Mises, de Kuiper, de Watson, et de Anderson-Darling est donnée en annexe.

Le tableau suivant montre les résultats obtenus par le calcul des différentes statistiques:

Groupe	D+	D-	W2	V	U2	A2
RNEP / CH / FCL / 20	2.147	2.122	.602	4.262	.552	5.529
/ 40	1.523	1.487	.419	3.009	.375	3.049
LCL / 20	1.859	1.781	.989	3.640	.783	5.605
/ 40	.674	.841	.157	1.515	.093	1.159
VIDE / 20	3.577	3.558	4.166	7.135	3.894	29.314
/ 40	3.815	3.848	4.153	7.663	3.973	24.511
FM / VIDE / 20	2.057	2.017	.837	4.068	.798	5.635
/ 40	2.388	2.440	1.259	4.828	1.194	7.408
RNRR / CH / FCL / 20	1.687	1.662	.713	3.349	.472	4.403
/ 40	3.703	3.661	3.647	7.364	2.117	17.155

La comparaison de ces résultats avec les tables correspondantes permet de rejeter ou d'accepter au niveau α l'hypothèse de normalité des résidus orthogonalisés (α % est la probabilité pour que H soit vraie alors qu'on la rejette)

Les blancs indiquent que H ne peut être acceptée à un niveau supérieur ou égal à 1 %

Les pourcentages sont les niveaux maximum auxquels on peut accepter H

Groupe	D	W2	V	U2	A2
RNEP / CH / FCL / 20		1 %			
/ 40	1 %	2.5 %			1 %
LCL / 20					
/ 40	15 %	10 %	5 %	10 %	10 %
VIDE / 20					
/ 40					
FM / VIDE / 20					
/ 40					
RNRR / CH / FCL / 20		1 %			
/ 40					

Ainsi l'hypothèse de normalité des résidus est-elle fortement contestable (mis à part les groupes LCL et FCL 40 pieds). Or c'était là une hypothèse de base que l'on a utilisée implicitement dans le calcul de différents tests ou intervalles de confiance. Il faut donc être très circonspect vis-à-vis de la fonction de coût obtenue.

Ce mauvais comportement des résidus est en fait confirmé par un examen attentif du graphe "résidus vs τ" qui semble être légèrement incurvé.

Deux explications peuvent être avancées:

1/ il manque une variable explicative, qui ne figurait pas dans le fichier d'observations de la compagnie (les variables que l'on a écartées étaient de façon évidente sans rapport avec la dépense)

2/ la fonction cherchée n'est pas linéaire. Aussi avons-nous tenté d'améliorer les résultats en transformant τ par une fonction g non linéaire.

τ = F(CONT, POID/CONT, KMDIS)
DEP = g(τ)

La fonction g appliquée à τ est une façon de contourner le problème des logiciels de régression qui n'acceptent dans des modèles non linéaire qu'une seule variable explicative. Mais les résultats obtenus sont encore décevants:

RNEP CH LCL 20'

τ = a + b ln(CONT) + c ln(POID/CONT) + d ln(KMDIS)
avec a, b, c, d coefficients obtenus par régression
p.23

DEP = a' exp(b' τ) + c' R^2= 98 %
a'= .223
b'=1.108
c'=41128
mais les résidus ne sont toujours pas normaux..

VI.CONCLUSION:

Parmi toutes les régressions faites, seule la régression qui conduit à une fonction du type

$$\text{Dépense} = A\,(\text{Distance})^{b}\,(\text{Poid Moyen})^{c}\,(\text{Nombre de conteneurs})^{d}$$

avec A, b, c, d caractéristiques du groupe concerné:

RNEP / CH / FCL
LCL et du nombre de pieds
VIDE (20 ou 40)
/ FM / VIDE
RNRR / CH / FCL

est acceptable, étant donné

le R^2
le graphe des résidus vs la variable expliquée
vs les variables explicatives
le graphe des résidus par rapport à la loi normale

Toutefois cette fonction ne conduit pas à des résidus distribués suivant une loi normale. Il nous faut donc choisir:

* soit nous la rejetons et nous cherchons une nouvelle fonction qui satisfasse tous les tests "académiques" (mais sous quelle forme? Nous nous heurtons aux problèmes de la non-linéarité multidimensionnelle et des variables à rajouter: comment les découvrir? D'autant que rien ne prouve que cette fonction parfaite existe);

* soit nous la conservons, connaissant ses limites, car elle semble très proche des données (résidus faibles, R^2 proche de 1).

Comme cette fonction s'inscrit dans le cadre plus large d'une planification à moyen terme de transport de conteneurs (prévision tactique et non réalité opérationnelle ou comptabilité) cette deuxième attitude semble la plus raisonnable. Deux "suites" sont dès lors à envisager pour cette étude:

* l'application de la démarche suivie dans cette étude et de ses résultats (forme de la fonction, identification des groupes) à l'estimation des coûts de transports par route (et non plus par rail);

* l'utilisation de la fonction trouvée pour créer un nouveau graphe, complet, où tous les trajets entre ports, dépôts et clients sont permis, et affectés d'un coût. Quelle est alors l'économie réalisée, par rapport à la première optimisation ?

BIBLIOGRAPHIE

T.G.Crainic, P.Dejax, L.Delorme (1986)
"Models for Multimode Multicommodity Location Problems with Interdepot Balancing Requirements"
CRT de l'Université de Montréal, publication #490A

P.Dejax, T.G.Crainic, L.Delorme, M.Bloise,
E.de Tocqueville, J.Hodgson (1988)
"Planification Tactique du transport terrestre des conteneurs vides"
Rapport de Recherche du LEIS de l'Ecole Centrale de Paris
et du CRT de l'Université de Montréal

W.J.Dixon (1983)
"BMDP Statistical Software Manual"
University of California Press

N.Draper et Harry Smith (1981)
"Applied Regression Analysis, Second Edition"
Ed Wiley-Interscience

S.R.Searle (1971)
"Linear Models"
Ed John Wiley & Sons

E.de Tocqueville (1987)
"Modélisation d'un système de transport terrestre de conteneurs vides"
Rapport de stage fin d'études ECP

M.Vallée (1979)
"Techniques Statistiques en Recherche Opérationnelle"
Notes du cours IFT-3670
Département d'informatique et de Recherche Opérationnelle
Université de Montréal

"Carte des zones pour le calcul des distances tarifaires"
éditée et imprimée par l'IGN
pour le Comité National Routier

ANNEXES

A.DESCRIPTION DES TESTS STATISTIQUES UTILISES

1. INTRODUCTION 2

2. ETUDE GENERALE 3
   a. R^2 Coefficient de corrélation multiple (coefficient de détermination) 3
   b. F-ratio 3
   c. t-test pour les coefficients de la régression 4
   d. Contribution to R^2 4
   e. F-to-remove & F-to-enter 5

3. EXAMEN DES RESIDUS 6
   a. Tests de normalité 6
   b. Tests de l'espérance nulle des résidus: . 7
   c. Tests d'indépendance des résidus: 8
   d. Homosédasticité de la variance des résidus 8

4. DETECTION DES OBSERVATIONS ABERRANTES OU PRéPONDéRANTES 9
   a. Distance de Mahalanobis. 9
   b. Distance de Cook 9

5. PRéCISION DES RéSULTATS 10
   a. Rapport (Ecart type des etimations / Moyenne) 10
   b. Intervalles de confiance (Présentation). . 10
   c. Intervalle de confiance pour les coefficients bi 11
   d. Intervalle de confiance pour la variable estimée y 12

B.FICHIER OBSFRD 13
   Variables 13
   Régions 14

C.TATI 15

D.REPARTITION DES MOUVEMENTS 16
   SUIVANT LE REGIONS ARRIVEE/DEPART
   SUIVANT LA TAILLE-TYPE

E.RESULTATS DES REGRESSIONS

1. INTRODUCTION

Toutes les régressions effectuées sont des régressions linéaires à une ou plusieurs variables.

Le logiciel BMDP fournit pour toutes les régressions un certain nombre d'indicateurs et tests usuels qui permettent d'en apprécier la validité.

Dans toute la suite on utilisera les notations suivantes:

Y variable à expliquer
r variable expliquée
x_i variables explicatives $0 \leq i \leq k$
e résidus

$$Y = b_0 + b_1 X_1 + b_2 X_2 + ... + b_k X_k + \epsilon$$
$$r = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$$
$$e = Y - r$$

N nombre d'observations
Y et r sont des vecteurs colomne de dimension [1,N]
y et r représentent une observation j ($1 \leq j \leq N$)
Xi sont des vecteurs colomne de dimension [1,N]
xij observation j de la variable i

b_i représente le coefficient "réel"
β_i est son estimation par la régression effectuée
ϵ est l'erreur, dont e est une estimation

2. ETUDE GENERALE

a. R^2 Coefficient de corrélation multiple (coefficient de détermination)

$$R^2 = \frac{\Sigma \text{ carrés dûs à la régression}}{\Sigma \text{ carrés}}$$

$$R^2 = \frac{\Sigma\ (\tau - y)^2}{\Sigma\ (y - y)^2} = \cos\ \theta$$

R^2 permet de savoir quel est le pourcentage des informations contenues dans Y, retenues dans Γ. Il peut également être interpreté comme la distance de Y à Γ, Γ étant finalement une projection de Y sur l'espace de dimension k+1 défini par les variables explicatives. Toutefois cette interprétation n'est valable que si la projection est orthogonale, c'est-à-dire les X_i indépendants. Cette hypothèse pouvant être vérifiée à l'aide de la matrice de corrélation des variables.

b. F-ratio

$$F = \frac{\Sigma \text{ carrés dûs à la régression } / \text{ k}}{\Sigma \text{ carrés des résidus } / \text{ (N-k-1)}}$$

$$F = \frac{\Sigma\ (\tau - y)^2\ /\ k}{\Sigma\ (y - y)^2\ /\ (N-k-1)} = \frac{R^2\ /\ k}{(1 - R^2)\ /\ (N-k-1)}$$

F permet de tester si le R^2 de l'échantillon est nul ou non, et déterminer si il est dû à une fluctuation sur les observations ou une erreur de mesure

En effet F doit suivre une loi F $(k, N-k-1, 1-\alpha)$

La mention de "significance" permet de connaître le meilleur pourcentage α pour le régression envisagée sans avoir besoin de se referrer à des tables de la loi F. de Fisher.

c. t-test pour les coefficients de la régression

Il teste l'hypothèse H_0: $b_i = 0$
vs H_1: $b_i = 0$

$$t = \frac{b_i}{s.e.(b_i)}$$ où s.e.(b_i) = écart type de b_i

t doit suivre une loi T (N-1,1-α/2)

De même que pour F, la mention de "significance" évite d'avoir à se re-errer à une table du Student en donnant directement le α optimal envisageable.

d. F-to-remove & F-to-enter

Ces statistiques sont utilisées dans les régressions pas-à-pas (Stepwise), du programme BMDP.

A chaque pas,

on retire la variable explicative dont le F-to-remove est le plus grand, pourvu qu'il soit supérieur à 4,

on rentre la variable explicative dont le F-to-enter est le plus grand, et à condition qu'il soit supérieur à 4.

On montre qu'ils sont rigoureusement équivalents au t-test puisque

$$F(1,\nu,1-\alpha) = \{ t(\nu,1-\alpha/2) \}^2$$

$$\text{F-to-remove} = \left| \frac{b}{s(b)} \right|^2$$

$$= \frac{SS(\text{rési. sans la variable}) - SS(\text{rési.})}{SS(\text{rési.}) \ / \ (N-k-1)}$$

$$\text{F-to-enter} = \frac{SS(\text{rési.}) - SS(\text{rési. avec la variable})}{SS(\text{rési.avec la var.}) \ / \ (N-k-2)}$$

avec SS somme des carrés

I. EXAMEN DES RÉSIDUS

Cette étude joue un rôle essentiel dans la validation du modèle et permet d'éliminer d'emblée de nombreux modèles dont le coefficient de détermination, R^2, était pourtant très élevé. De plus l'examen des résidus peut donner des indications sur la façon de corriger le modèle pour obtenir de meilleurs résultats.

Comme on a fait l'hypothèse d'un modèle du type

$$y = Xb + e \qquad \text{avec } e \in N(0, \sigma^2 I)$$

les résidus ϵ doivent vérifier un certain nombre de propriétés pour que l'on puisse valider le modèle. En effet ils constituent une bonne approximation des erreurs e puisque

$$\tau = X\beta \quad \text{où } \beta \text{ est une estimation de } b$$

$$\text{et donc } \epsilon = X(b-\beta) + e$$

Les quatre propriétés à contrôler lors d'une étude de résidus sont:

1/ Normalité des résidus
2/ Espérance nulle des résidus
3/ Indépendance des résidus
4/ Homosédasticité des résidus

a. Tests de normalité

* Graphe des résidus par rapport à une loi normale (Normal probability plot of residuals):

* Graphe des résidus par rapport à une loi normale corrigé de l'effet de tendance (Detrended normal probability plot of residuals)

* Histogramme des résidus studentisés: ils sont définis comme $(y - \tau)/s.e.e.$ où s.e.e. est l'écart type de la variable estimée (standard error of estimated) c'est-à-dire

$$\sqrt{\Sigma (y - \tau)^2 / (N-k)} = s.e.e.$$

* Tests sur les résidus orthogonalisés: plusieurs tests, décrits en détail par Marcel Vallée dans "Techniques Statistiques en Recherche Opérationnelle" (Notes de cours) ont été utilisé.

Statistiques de Kolmogorov-Smirnov D+, D-, D = max (D+,D-)	T* = T √N
Statistique W2 de Cramer-von Mises	T* = T
Statistique V de Kuiper	T* = T √N
Statistique U2 de Watson	T* = T
Statistique A2 de Anderson-Darling	T* = T

Pour faire un test utilisant l'une de ces statistiques (T) on calcule sa forme modifiée (T*). Pour chacune, des tables ont été établies qui amènent à rejeter l'hypothèse H "les résidus orthogonalisés suivent une loi normale $N(0,\sigma^2)$" au niveau α si et seulement si T* > T critique.

b. Tests de l'espérance nulle des résidus:

A partir du moment où le modèle possède une constante il est évident que cette propriété sera vérifiée. Aussi, pour valider un modèle est-il nécessaire de faire le calcul de la moyenne sur les résidus orthogonalisés.

Ceci nous amène à rejeter H: "la moyenne des résidus orthogonalisés est nulle" au niveau α si et seulement si

$$\left| \frac{X\text{-bar}}{s \,/\, \sqrt{N}} \right| \geq t\,(N-1, 1-\alpha/2)$$

où X-bar = moyenne des résidus orthogonalisés X
$s^2 = \Sigma\,(X - X\text{-bar})^2 \,/\, (N-1)$

La comparaison de cette valeur avec des valeurs tabulées fournit un encadrement de α, et donc un niveau maximal d'acceptation de l'hypothèse H.

c. Tests d'indépendance des résidus:

Un test intéressant est souvent celui de Durbin-Watson qui permet de s'assurer que les ϵ ne sont pas de la forme $\epsilon_j = \mu \epsilon_{j-1} + f_j$ où les f_i sont indépendamment et identiquement distribués selon une loi $N(0, \sigma_2)$. Dans ces conditions il est évident que les ϵ_j seraient corrélés entre eux.

D'où la statistique de Durbin-Watson définie par

$$D = \Sigma (\epsilon_j - \epsilon_{j-1})^2 / \Sigma \epsilon_j^2$$

Pour tester l'hypothèse H ($\mu = 0$), il suffit, après avoir calculé D, de se reporter à des tables donnant pour différents niveaux de confiance, les valeurs de d_L et d_U telles que

si $D > d_U$ on accepte H
si $d_U \geq D \geq d_L$ on ne peut conclure
si $d_L > D$ on rejette H

Ces tables prennent en compte α (level of significance)
k
N

Cependant ce test est très sensible à la façon dont sont ordonnées les observations (c'est-à-dire, ici, par ordre croissant des cantons de départ puis des cantons d'arrivée). Il se pourrait que les ϵ soient corrélés suivant un ordre différents: le test D-W ne le détecterait pas.

d. Homosédasticité de la variance des résidus

Les graphes e_j vs y_j, r_j, x_{ij} permettent de l'étudier: un "bon" modèle devant donner des graphes ayant tous leurs points à l'intérieur d'une bande horizontale centrée sur 0.

4. DETECTION DES OBSERVATIONS ABERRANTES OU PRÉPONDÉRANTES

Elles sont identifiées à l'aide des distances mentionnées par le programme BMDP9R:

a. Distance de Mahalanobis

$$(N-1)\ (X-\bar{X})\ C^{-1}\ (X-\bar{X})'$$

où X est la matrice des observations
C est la matrice de variance-covariance

Elle représente la distance de chaque observation à la moyenne de toutes les observations utilisées dans la régression. Une distance importante est significative d'un "outlier" (donnée hors-limite).

b. Distance de Cook

$$\frac{r^2}{(k+1)\ RMS}$$

où r est le résidu pour l'observation si on la retire de la régression (deleted residual)
RMS residual mean square

Elle mesure la variation apportée dans les coefficients de la régression quand on retire l'observation considérée. Elle permet donc de repérer les points les plus influents et de vérifier qu'ils ne sont pas aberrants.

5. PRÉCISION DES RÉSULTATS

a. Rapport (Ecart type des estimations / Moyenne)

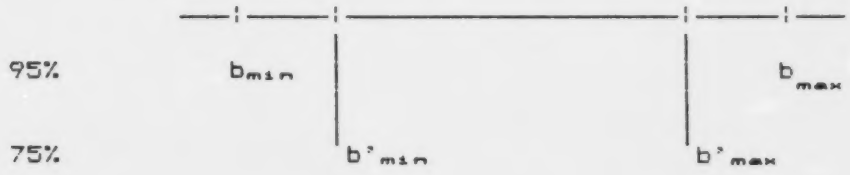
Ce rapport constitue un moyen très rapide de se rendre compte de la valeur des estimations en donnant un ordre de grandeur de l'erreur commise (l'écart-type peut être interprété comme une "erreur moyenne", que l'on rapporte ainsi à la moyenne de la variable sur laquelle elle est commise).

b. Intervalles de confiance

Afin de savoir quelle est la précision de l'équation obtenue on peut donner pour les coefficients de régression et pour la variable estimée des "intervalles de confiance" à $(1-\alpha)\%$

Dire que b appartient à un intervalle $[b_{min}, b_{max}]$ avec une confiance de $(1-\alpha)\%$ signifie que la probabilité pour b d'être hors de cet intervalle est inférieure ou égale à $\alpha\%$

Ainsi quand on réduit α la probabilité d'erreur diminue mais la largeur donc la précision également:



2. Intervalle de confiance pour les coefficients bi

Comme on fait l'hypothèse d'une loi normale $N(0,\sigma^2)$ pour les résidus $y - \pi$, cela implique, pour un niveau de confiance de $(1-\alpha)\%$

bi' compris entre $\beta i \pm t(N-2 , 1-\alpha/2)\ \sigma(\beta i)$

où $\sigma(\beta i)$ est l'écart-type de βi
et βi valeur estimée de bi

Les tables de la loi t indiquent

à 80% N ≥ 120 $t(N-2 , 1-\alpha/2) = 1.29$
à 95% N ≥ 120 $t(N-2 , 1-\alpha/2) = 1.98$
à 99% N ≥ 120 $t(N-2 , 1-\alpha/2) = 2.62$

Pour déterminer les intervalles de confiance de cette étude on a pris t = 2

d. Intervalle de confiance pour la variable estimée "

On a pour l'observation j

$$y - b_o = b_1 x_{1j} + b_2 x_{2j} + ... + b_k x_{kj} + \epsilon$$
$$\tau - \beta_o = \beta_1 x_{1j} + \beta_2 x_{2j} + ... + \beta_k x_{kj}$$

La variance de $\tau-\beta_o$ est

$$V(\tau-\beta_o) = V(\beta_1 x_{1j} + \beta_2 x_{2j} + ... + \beta_k x_{kj})$$

$$= V(\beta_1)\, x_{1j} + ... + V(\beta_k)\, x_{kj} + 2\, x_{1j}\, x_{2j}\, Cov(\beta_1,\beta_2) + ...$$

Matriciellement on peut réecrire cette expression sous la forme:

$$V(\tau) = \sigma^2 \; [\, x1j \; ... \; xij \,] \; C \begin{vmatrix} x1j \\ \\ xkj \end{vmatrix} = \sigma^2 \; X'_j C X_j$$

où C est la matrice de variance des coefficients de la régression.

Ceci donne alors pour intervalle de confiance à $(1-\alpha)\%$

$$\tau-\beta_o \pm t(N-k-1 , 1-\alpha/2)\; s \sqrt{X'_j C X_j}$$

où s est le carré moyen des résidus

La encore on a toujours pris t = 2 pour déterminer les intervalles de confiance.

Puis pour tenir compte de la constante on a majoré par l'écart type de β_o. Finalement cela donne

$$\tau \pm 2 \; [\; \sigma(\beta_o) + s \sqrt{X'CX} \;]$$

B.FICHIER OBSFRD

Variables

	Columns begin	end	Name & Description
A3	1	3	CANTD (canton de depart)
I2	4	5	REGD (région de départ)
A3	6	8	CANTA (canton d'arrivée)
I2	9	10	REGC (région d'arrivée)
A3	11	13	CANTC (canton de cammionage)
I2	14	15	REGC (region de cammionage)
I4	16	19	EXPE (nombre d'expéditions)
I4	20	23	KMCGM (distance observée)
I4	24	27	CONT (nombre de conteneurs)
I2	28	29	TATI (type de conteneur)
I4	30	33	POID (poid total pour ce transport)
I2	34	35	MODE | 3 RNEP Rail National Embranchement Particulier | 7 RNRR Rail National combiné Rail Route
A2	36	37	Code mouvement
I1	38	38	RE (régime) | 1 CH Carrier Haulage | 2 FM Formule Mixte
I1	39	39	Code F | 1 FCL Full Container Load (Plein) | 2 LCL Less than Container Load (Incomplet) | 3 VIDE
I8	40	47	DEP (dépense)
I4	48	51	KMDIS (distance obtenue à partir du distancier rail)

Le Carrier Haulage CH correspond à un transport aux frais du transporteur, par opposition au Merchant Haulage MH (aux frais du client) ou à la Formule Mixte FM (la compagnie amène le conteneur à ses frais chez le client, celui-ci paie le transport une fois rempli).

Ce fichier a été trié par ordre croissant de
1/ Canton de départ
2/ Canton d'arrivée

Le code utilisé pour les numéros de cantons est le code administratif français:
les 2 premiers chiffres indiquent le département
le 3ème est propre au canton

Pour définir et numéroter les régions on a également suivi les découpages en région administratives

Régions	Départements
Région 1	
Nord - Pas de Calais	59 62
Picardie	02 60 80
Ile de France	75 77 78 91 92 93 94 95
Haute Normandie	27 76
Région 2	
Centre	18 28 36 37 41 45
Région 3	
Basse Normandie	14 50 61
Bretagne	22 29 35 56
Pays de Loire	44 49 53 72 85
Région 4	
Poitou Charentes	16 17 79 86
Limousin	19 23 87
Aquitaine	24 33 40 47 64
Midi Pyrénées	09 12 31 32 46 65 81 82
Région 5	
Languedoc Roussillon	11 30 34 48 66
Provence Côte d'Azur	04 05 06 13 83 84
Région 6	
Auvergne	03 15 43 63
Rhônes Alpes	01 07 26 38 42 69 73 74
Région 7	
Champagne Ardennes	08 10 51 52
Lorraine	54 55 57 88
Alsace	67 68
Franche Comté	25 39 70 90
Bourgogne	21 58 71 89

Région hors de France		
Belgique	(B)	8
Hollande	(N)	9
Luxembourg	(L)	10
RFA	(D)	11

C.TATI

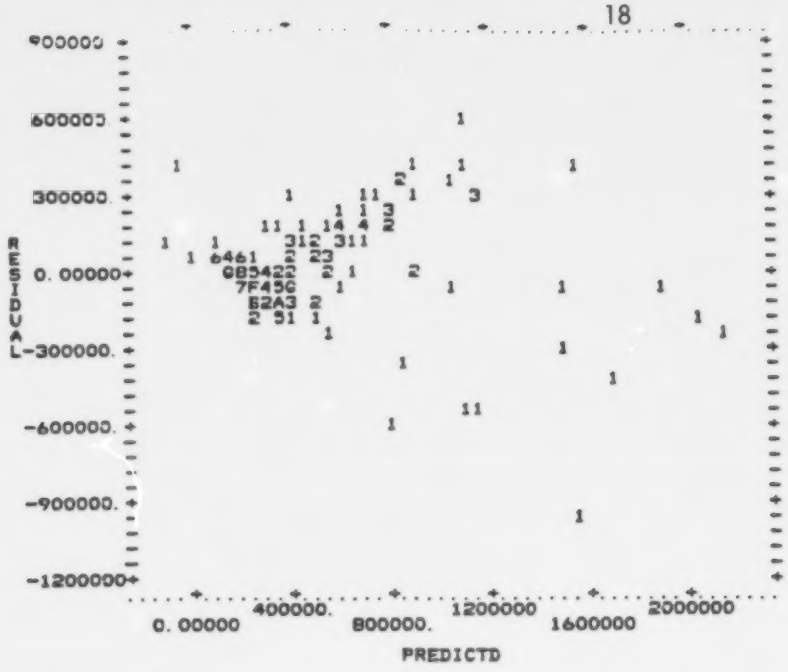
Description des codes "TATI" correspodant à la taille et au type des conteneurs transportés:

CODE TATI	CODE TAILLE	CODE TYPE	DESIGNATIONS & OBSERVATIONS
1	20	00	Conteneur 9m3
2	20	40	20' Usage général, toute hauteur
3	20	51	20' Open top
4	20	99	Divers 20'
5	22	32	20' Frigo
6	22	49	20' Isotherme
7	40	51	40' Open top
8	40	99	Divers 40'
9	42	40	40' Isotherme (sauf ant francaises)
10	42	49	40' Isotherme (ligne ant francaises)
11	43	00	40' Usage Général
12	43	32	40' Frigo

D.REPARTITION DES MOUVEMENTS
SUIVANT LES REGIONS DEPART/ARRIVEE
ET SUIVANT LES TAILLES-TYPES

REGA	REGD	TATI 1	2	3	4	5	6	20' TOT	TOT
1	1	50		1				51	
	5	62						62	113
2	1						44	44	
	2	14		1			1	16	60
3	3	9					2	11	
	5	4						4	15
4	1	6					68	74	
	4	76		2			1	79	
	5	123						123	276
5	1	8					185	193	
	4	1					1	2	
	5	22			2		10	34	229
6	1	10		1			156	167	
	5	5						5	
	6	219		3			17	239	
	7	1						1	412
7	1	14					12	26	
	4	1						1	
	5	127						127	
	7	369	1	7	1		10	388	542

RES4	RES	TATI 7	8	9	10	11	12	40' TOT	TOT
1	1					7		7	
	5					22		22	29
2	1				5			5	
	2					1		1	6
3	3					2		2	2
4	1				25	1		26	
	4	1			1	20		22	
	5					20		20	68
5	1				20	1		21	
	5	2	1		6	28	1	38	59
6	1				37	3		40	
	5					2		2	
	6	13			11	109		133	175
7	5					103		103	
	7				6	149	3	158	261

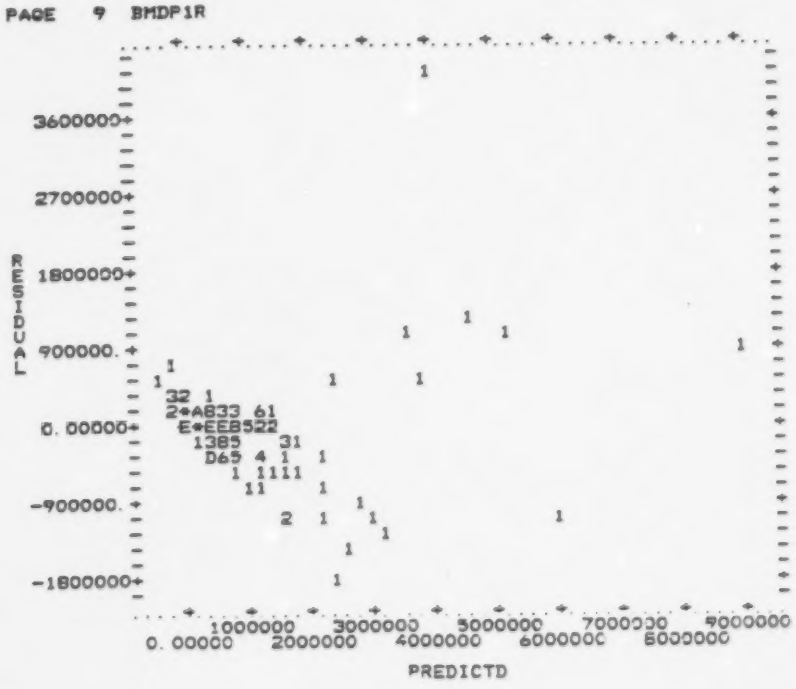


Co/FCL/ANEP/ Tati=1
Riga=1

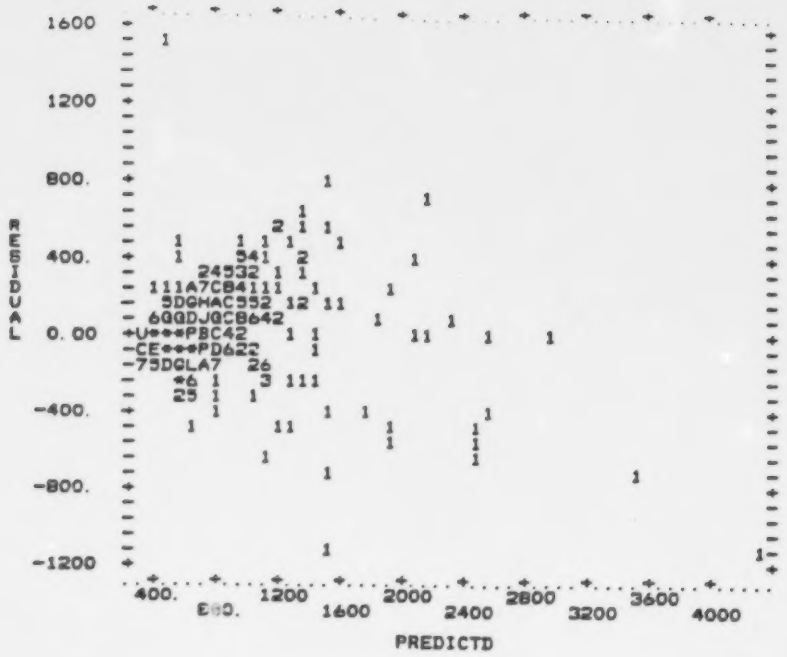
Kmdic < 330 km
Paid < 6.10^3
Dep < 8.10^6

Dep = a + b Kmdic + c Cont + d Paid

R^2: 81 %



CH/FCL/ANEP / Tati = 1
Riga = 1

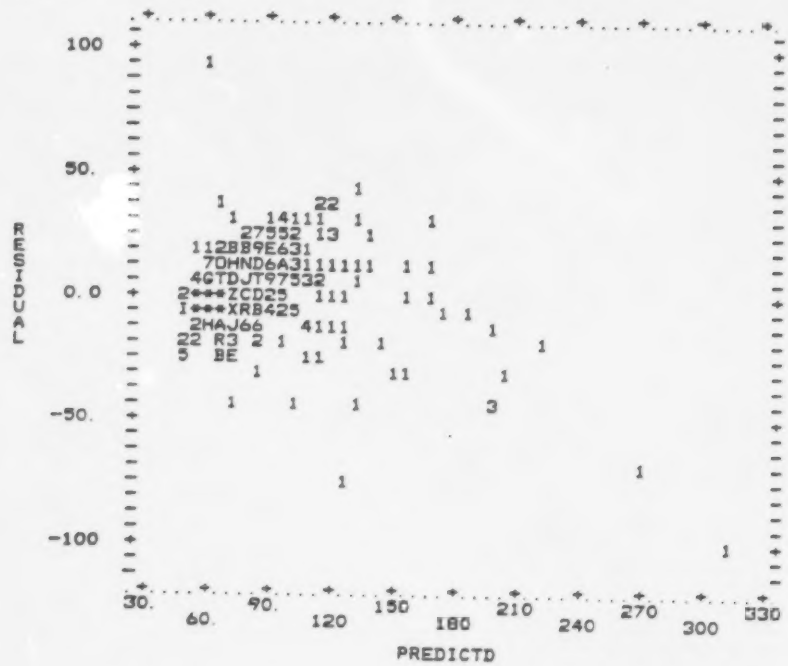


CH/FCL/RNEP/Tati=1
Rega=1

$\sqrt{Dep} = a + b\ Kmolis + c\ Const + d\ Paid$

$R^2 = 79\%$

PAGE 9 BMDP1R



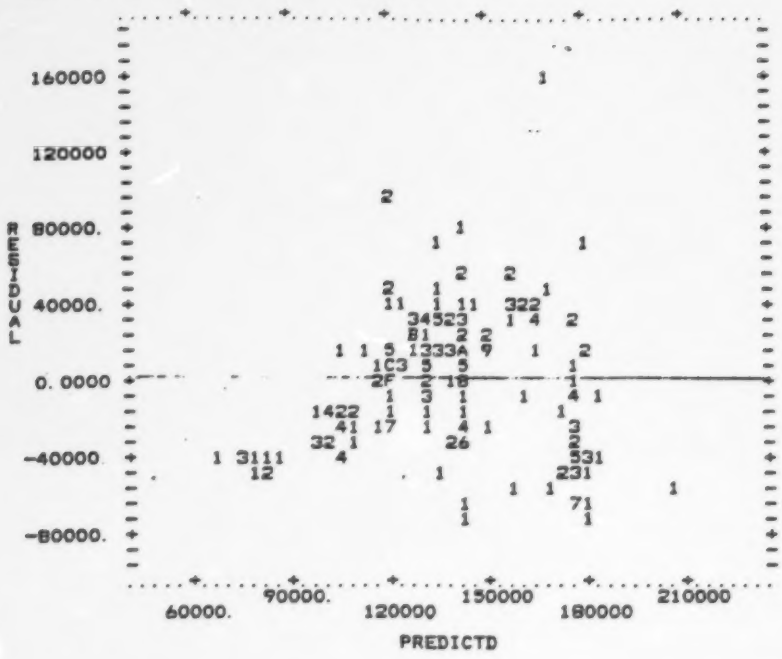
CH/FCL/RNEP/Tati=1
Rega=1

$\sqrt[3]{Dep} = a + c\ Kmolis + c\ Const + d\ Paid$

$R^2 = 75\%$

PAGE 5 BMDP1R REG DEPTRA/CONT 1

CH/FCL/RNEP / Tate = 1
Rega = 1

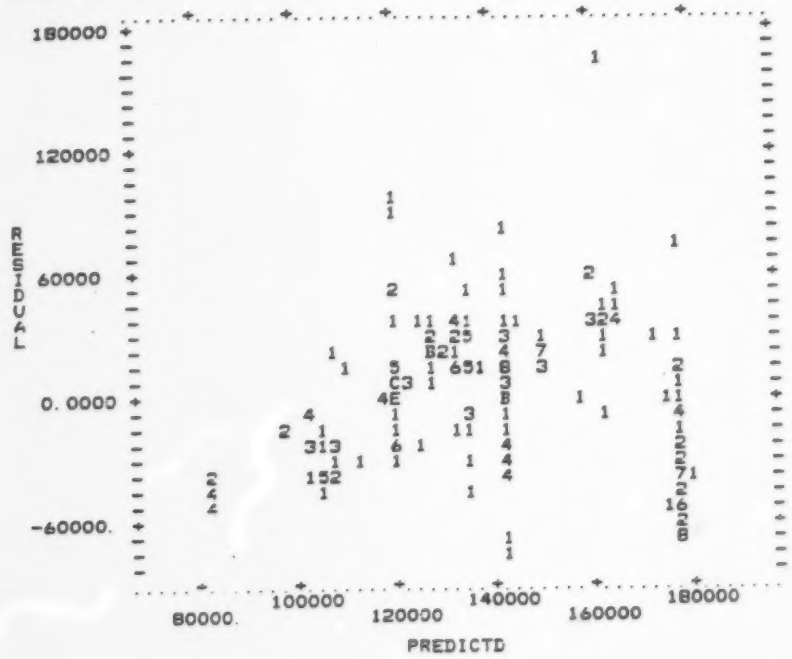


Dep / Cont : a + b Vmdic + c Cont + d Paid

R^2 : 37%

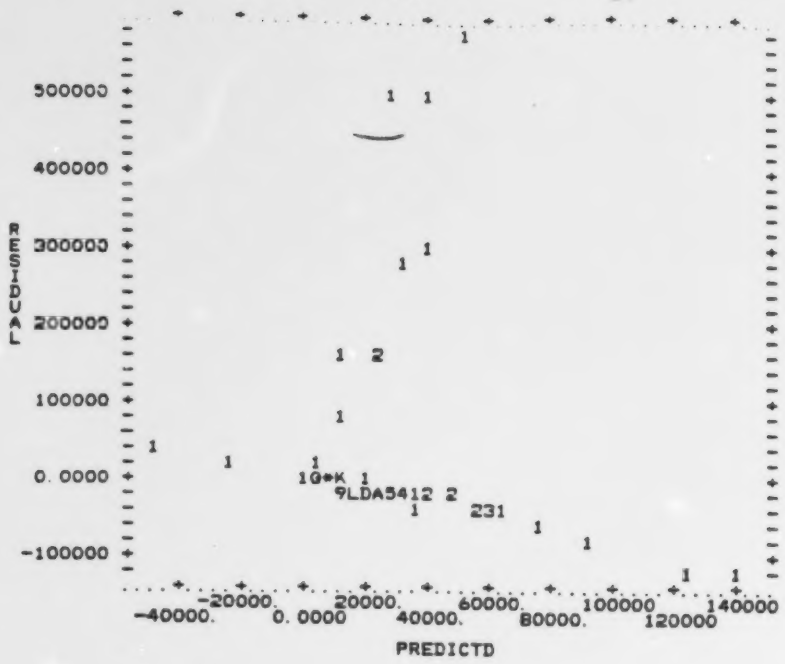
PAGE 10 BMDP1R

CH/FCL/RNEP / Tate = 1
Rega = 1



Dec / Cont : a + b Vmoc + c Cont + d Fac / Cont

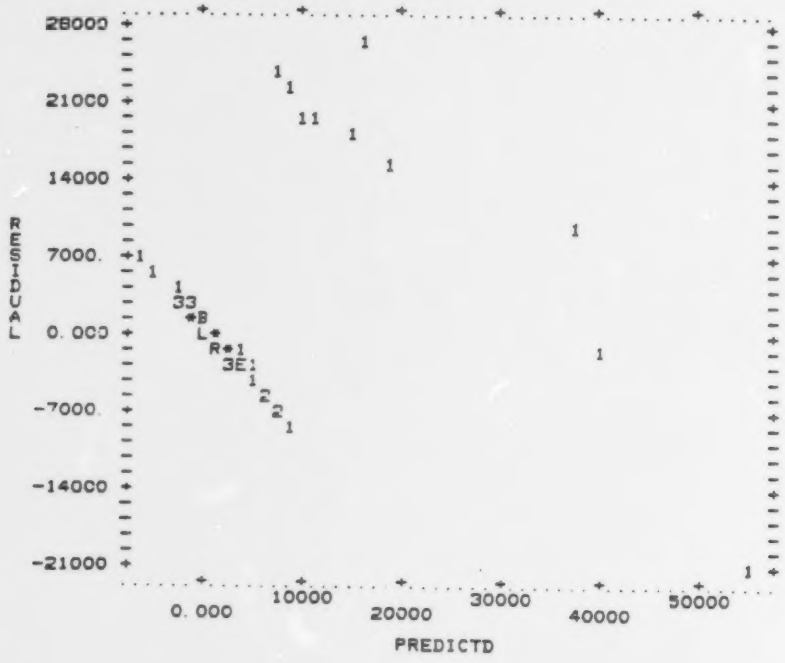
CH/FCL/RNEP ; Tati = 1
Rega = 1



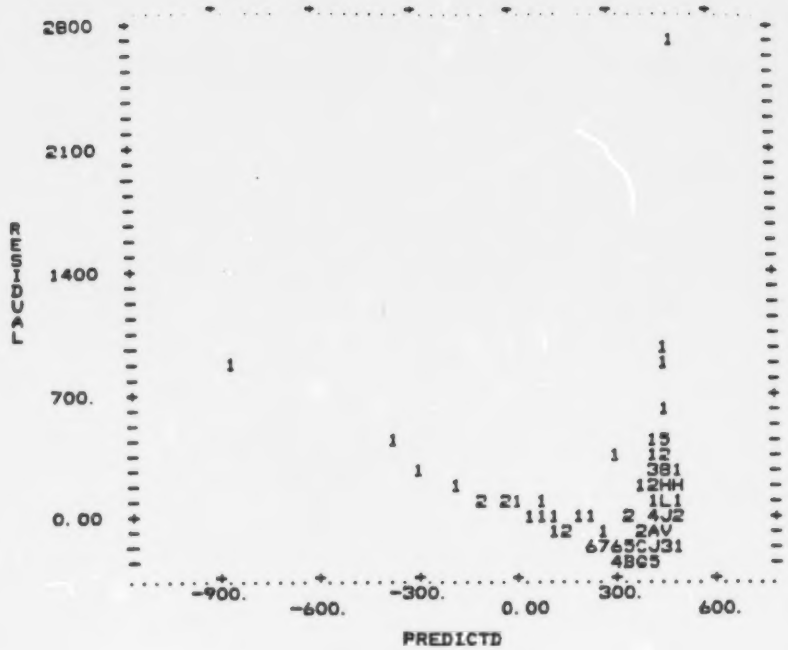
$\frac{Dep}{Kmdis} = a + b\ Kmdis + c\ Cont + d\ Paid$

PAGE 5 BMDP1R REG DEPTRA/(CONT*KMS) 1

CH/FCL/RNEP/ Tati = 1
Rega = 1



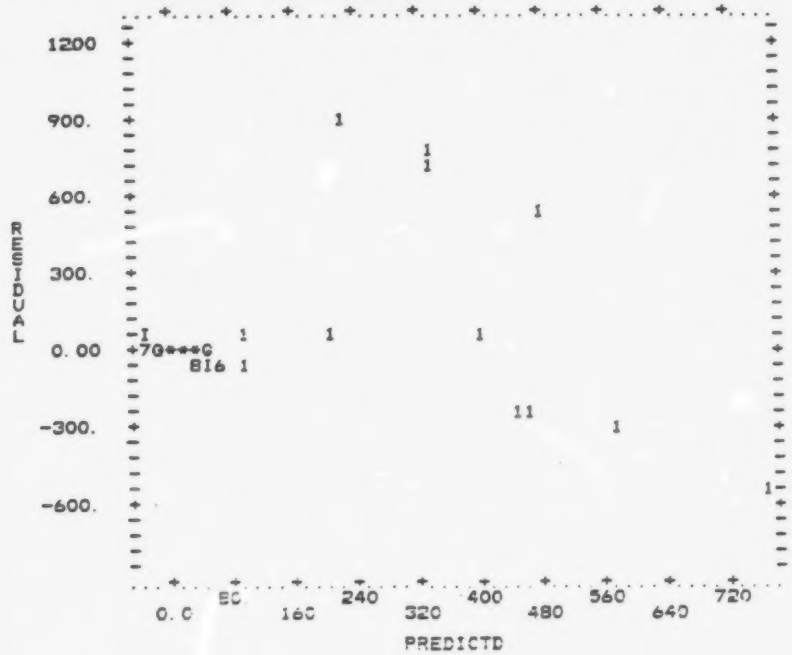
$\frac{Dep}{Cont\ Imois} = a + c\ Imois + c\ Cont + d\ Ficb + e\ \frac{Pilc}{\ldots} + \ldots \frac{Ire}{Imois}$



CH/FCL/RNEP/Tati=1
Rega=1

$\frac{1}{Dep} = a + b\ Vmoy + c\ Cout + d\ Poid$

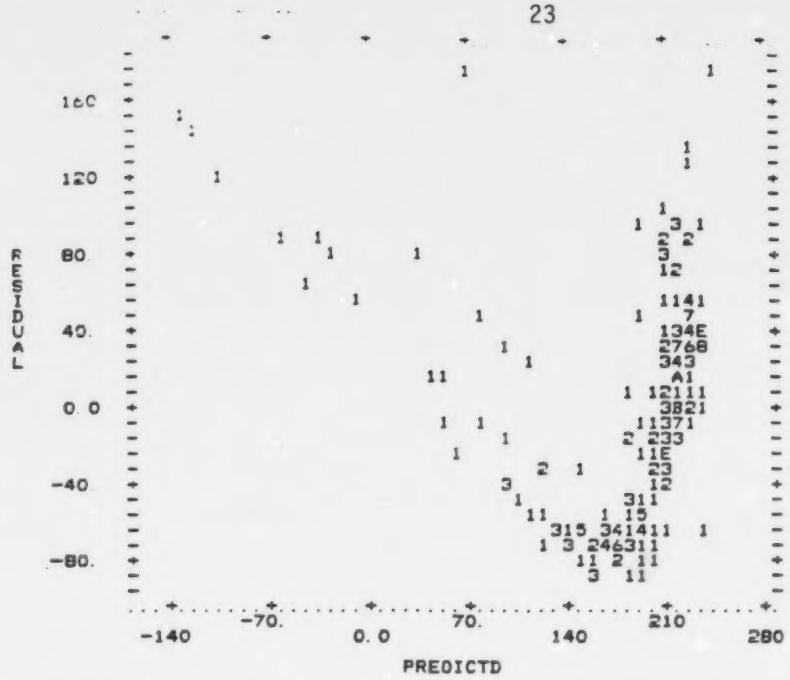
PAGE 8 BMDP2R REG STEPWISE DEPTRA/(POID*KMS) 1



CH/FCL/RNEP/Tati=1
Rega=1

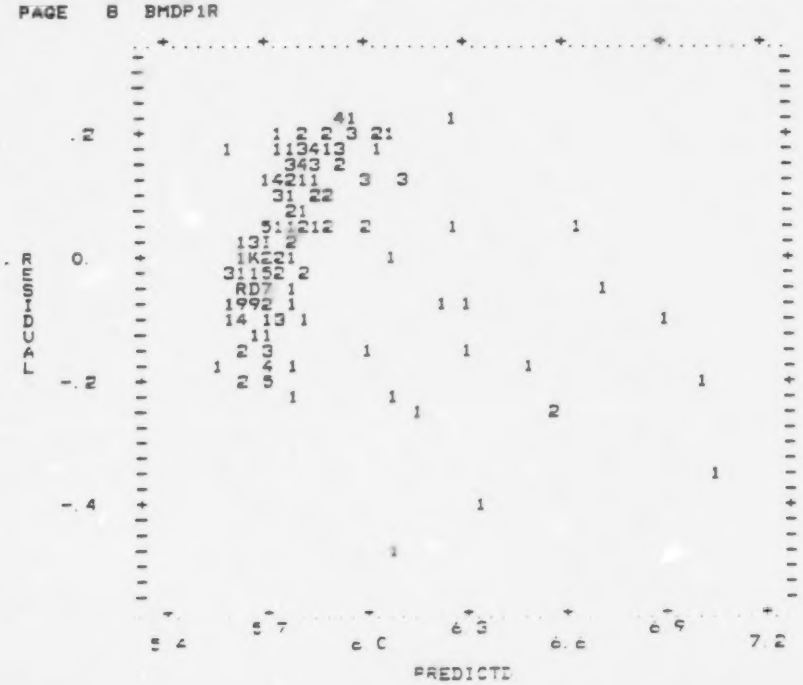
$\frac{Dep}{Vmoy\ Poid} = a + b\ Vmoy + c\ \frac{Cout}{Poid} + d\ \frac{Cout}{Poid}$

Remarque: Cout et Poid ne sont pas retenus dans la régression pas à pas

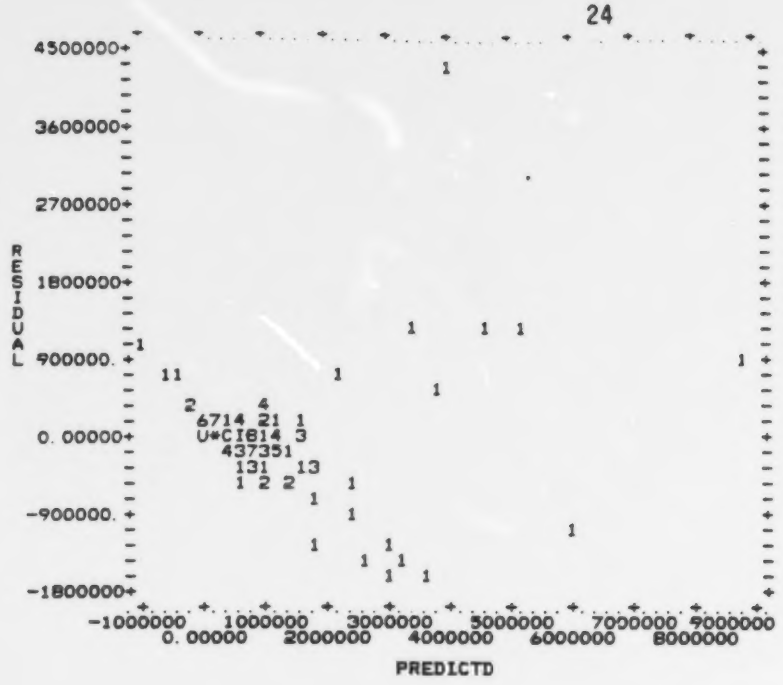


FCL/CH/RNEP / Tate = 1 Rega = 1 $\quad \frac{1}{\text{Dep + cte}} = a + b\,\text{Kmdia} + c\,\text{Cont} + d\,\text{Prod}$

PAGE 8 BMDP1R



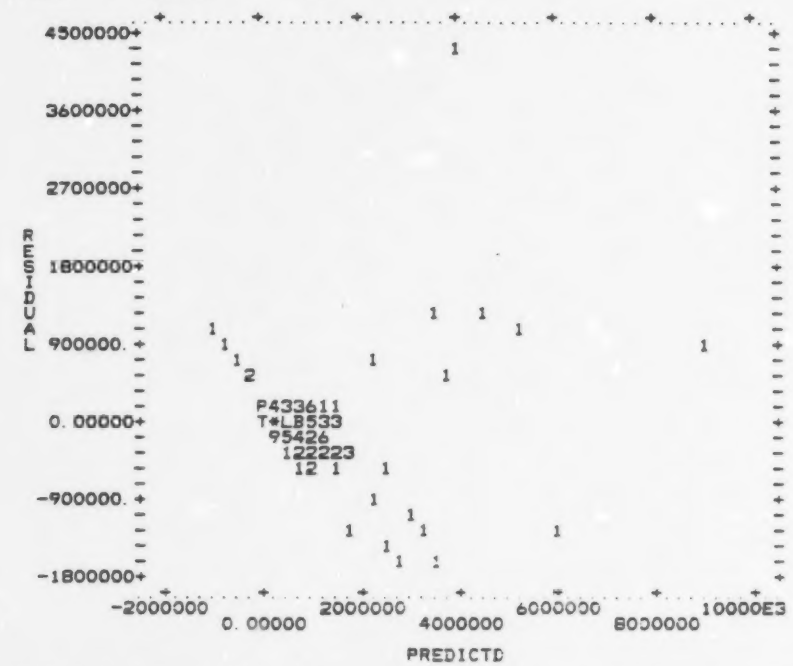
FCL/CH/RNEP / Tate = 1 RegA = 1 $\quad \ln(\text{Dep + cte}) = a + b\,\text{Kmdia} + c\,\text{Cont} + d\,\text{Prod}$



FCL / CH / RNEP / Tati = 1
Rega = 1

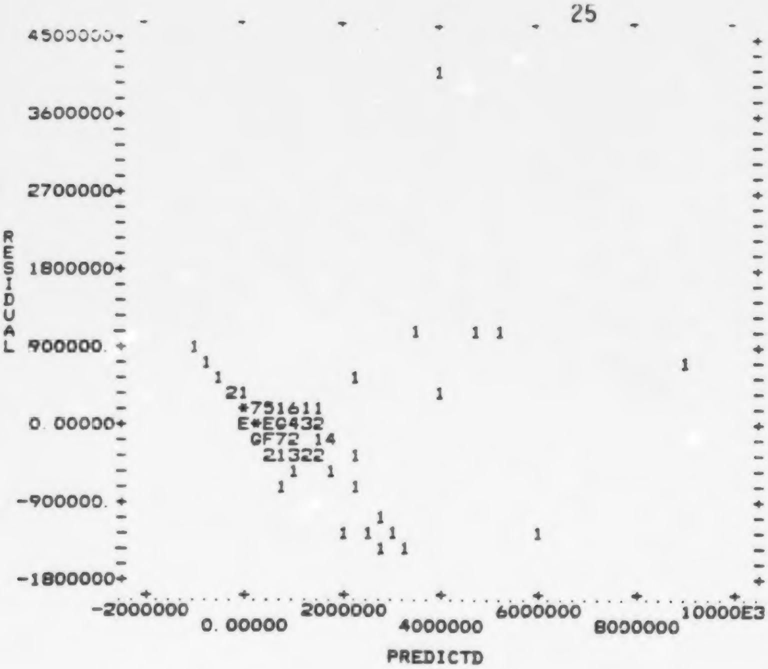
$Dep = a + b\ Cost + c\ Paid + d\sqrt{\ln(Klmdu)}$

PAGE 9 BMDP1R



FCL / CH / RNEP / Tati = 1
Rega = 1

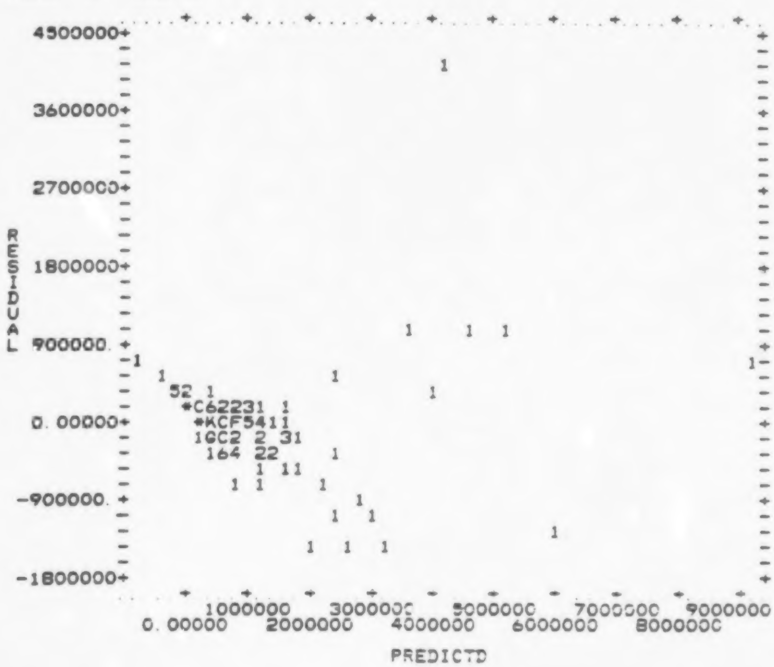
$Dep = a + b\ Cost + c\ Paid + d\ \ln(Klmdes)$



FCL/CH/RNEP / Tati = 1 Dep: a + b Cont + c Pord + d ln²(Kmdis)

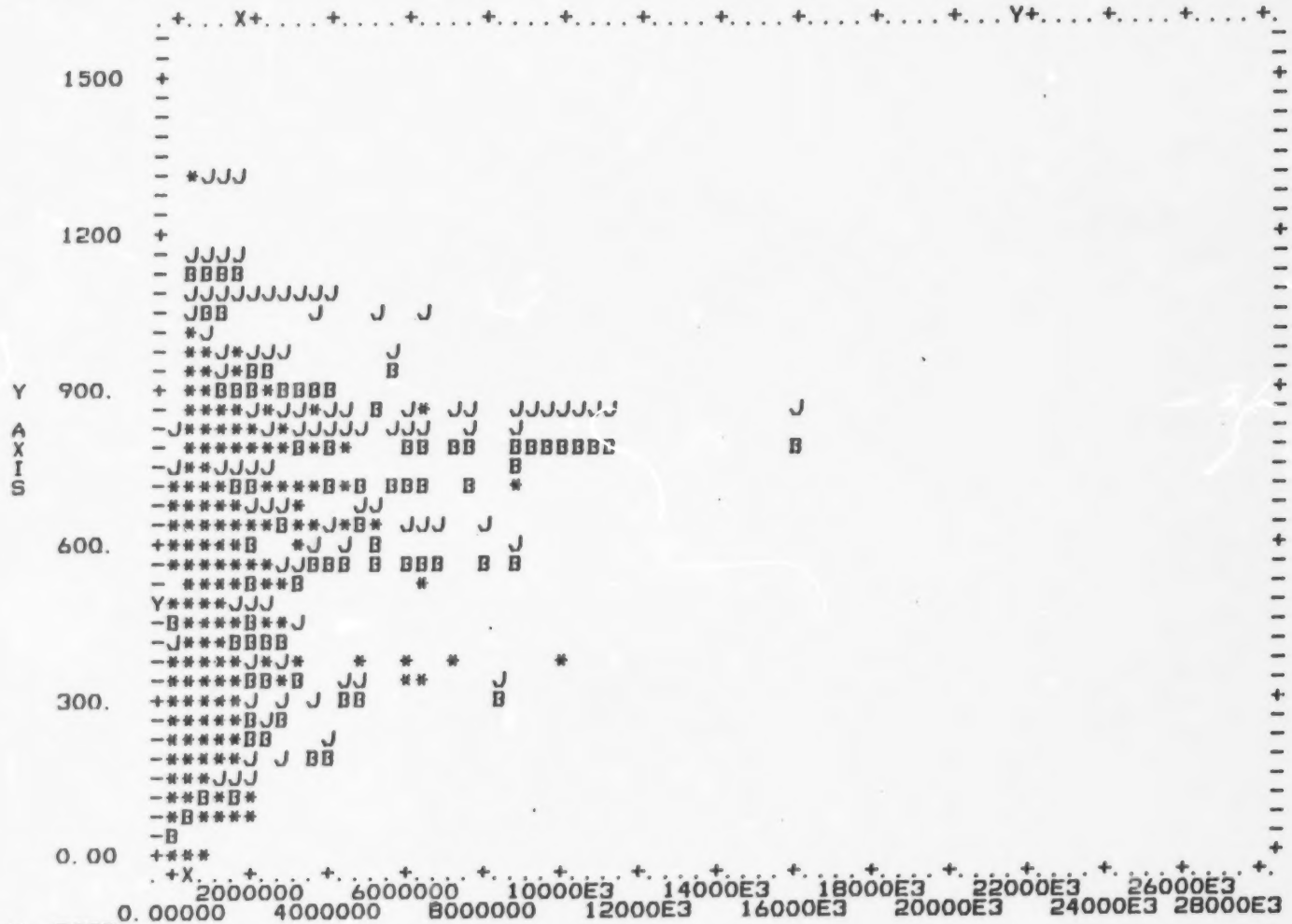
Rega = 1

PAGE 17 BMDP1R



FCL/CH/RNEP/Tati Dep: a + b Cont + c Pord + d ln³(Kmdis)

Rega = 1



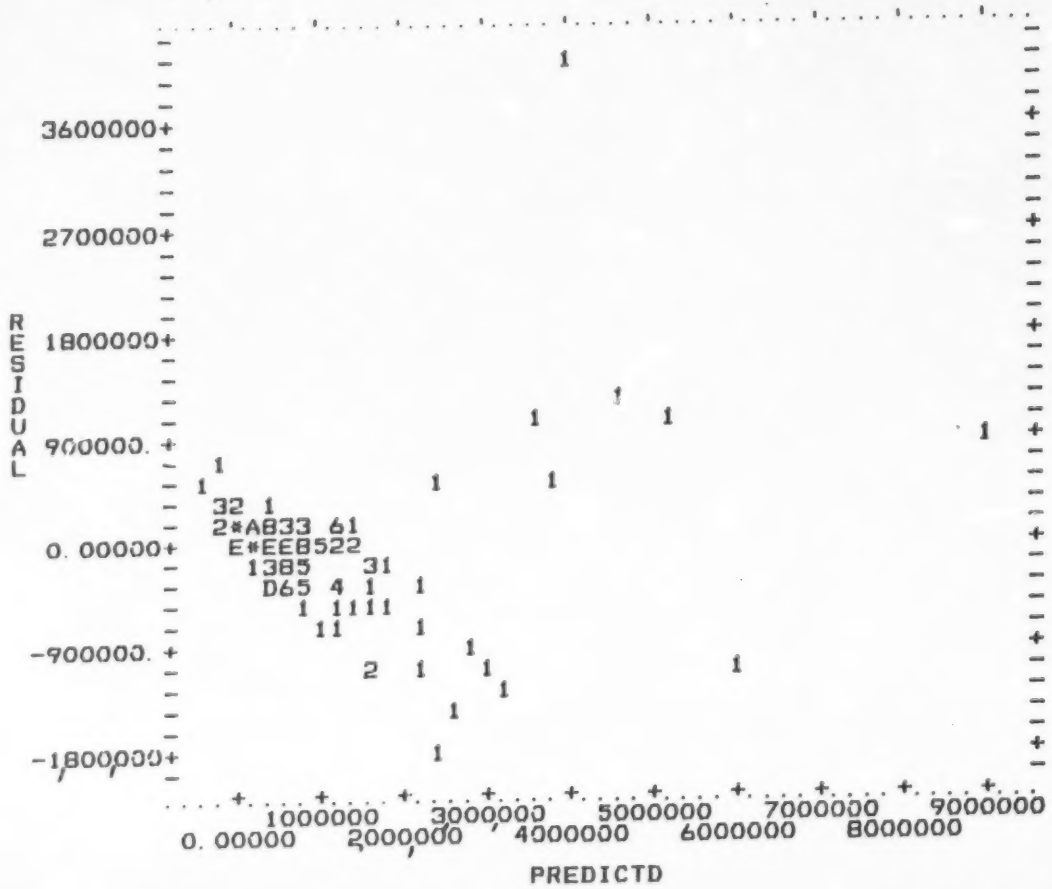
N =7092
R = .2027
P(R) 0.000

DEP

	MEAN	ST. DEV.	REGRESSION LINE	RES. MS.
X	745198	1063E3	X= 800.53*Y+ 307789	1084E9
Y	546.40	269.29	Y= 513E-7*X+ 508.14	69547.

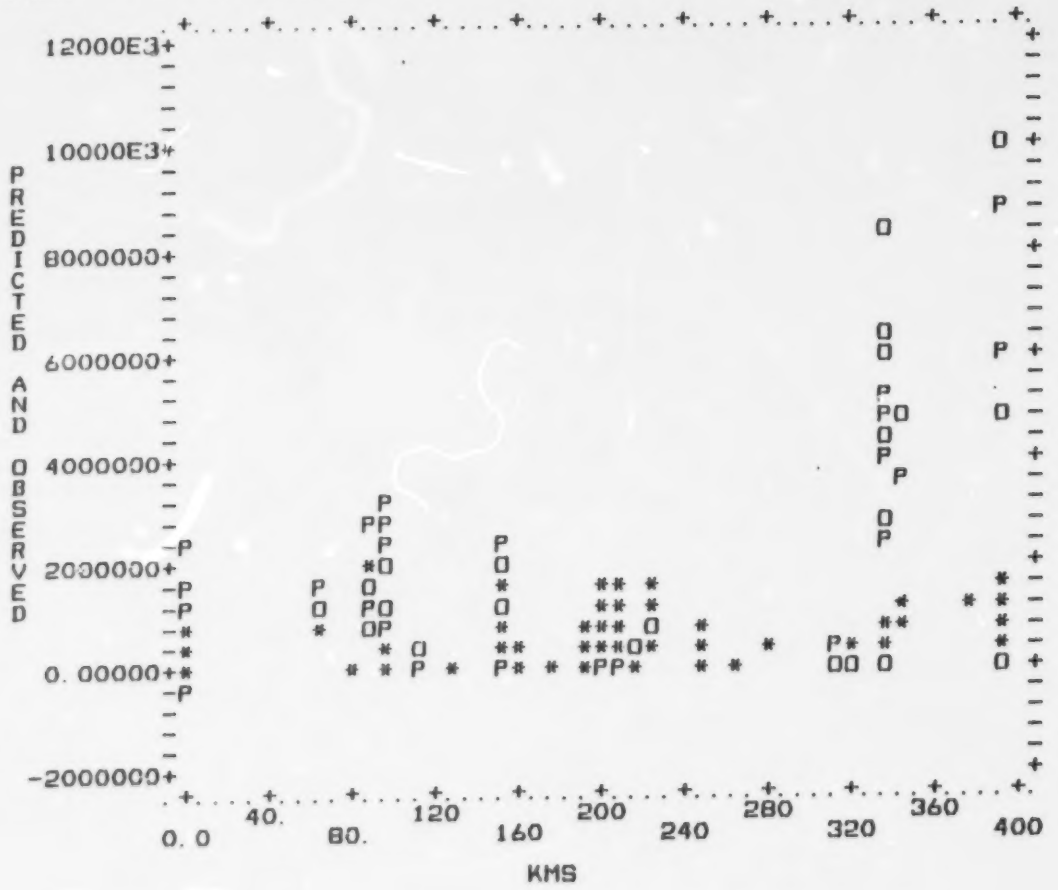
XVAR = 9 DEP VERSUS YVAR = 2 KMCGM SYMBOL=B
XVAR = 9 DEP VERSUS YVAR = 10 KMDIS SYMBOL=J

RNEP / FCL / CM.
Km < 1600



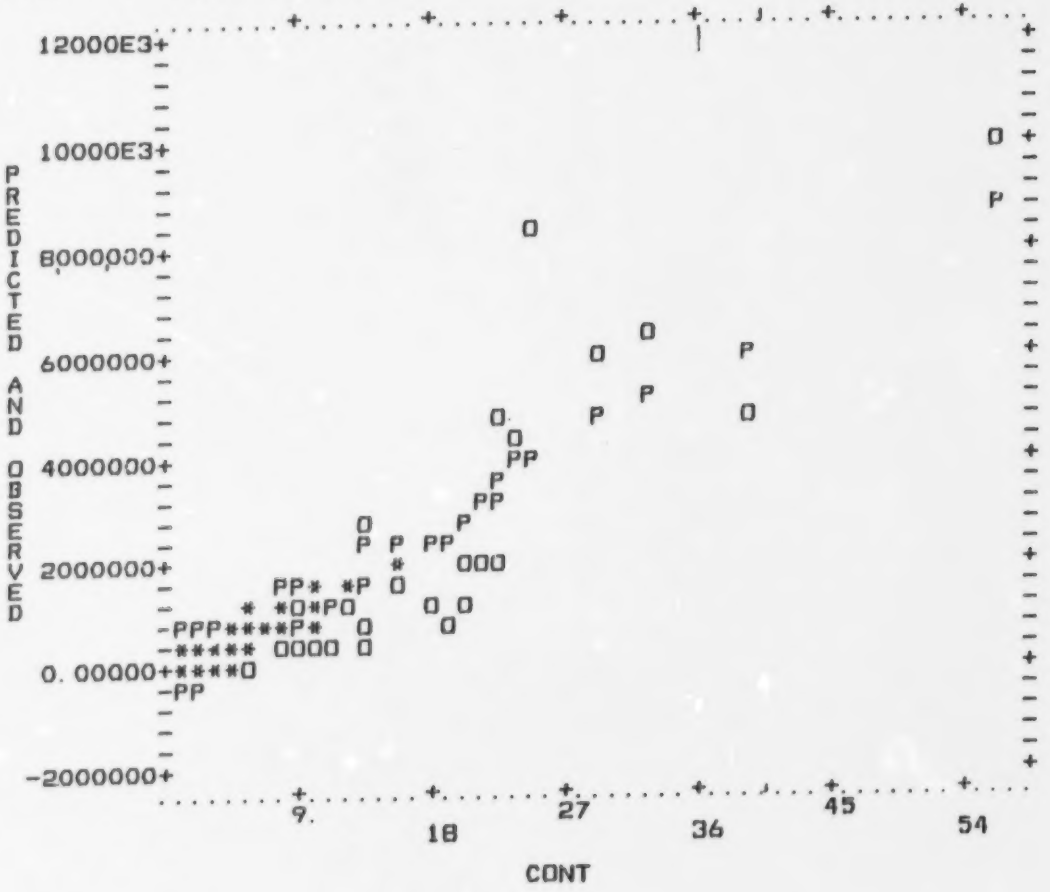
RNEP / FCL / CH / Tatê=1 / RegA=1

Regr. Dep fn linéaire de Km dis
Cent
Pond

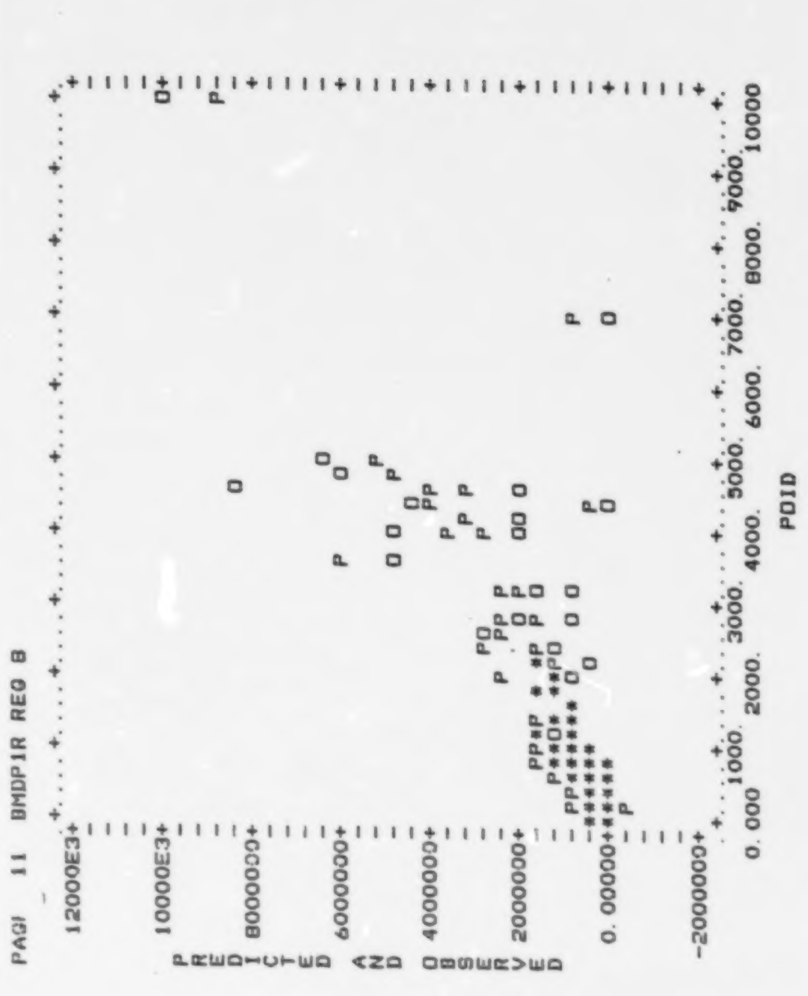


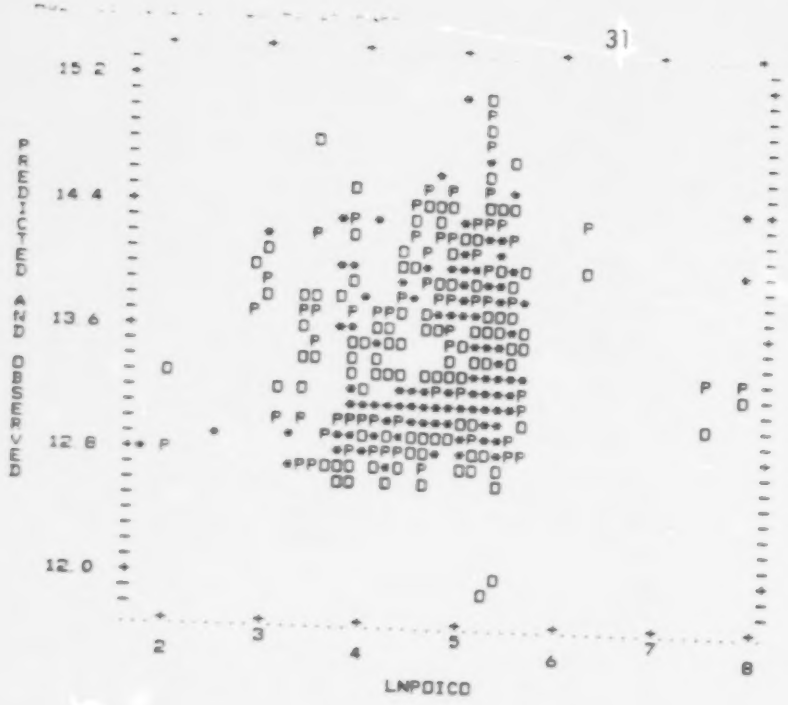
RNEP FCL CH Tate = 1 Rega = 1

Regr Dep p" lin de Vendus
Cont
Ord

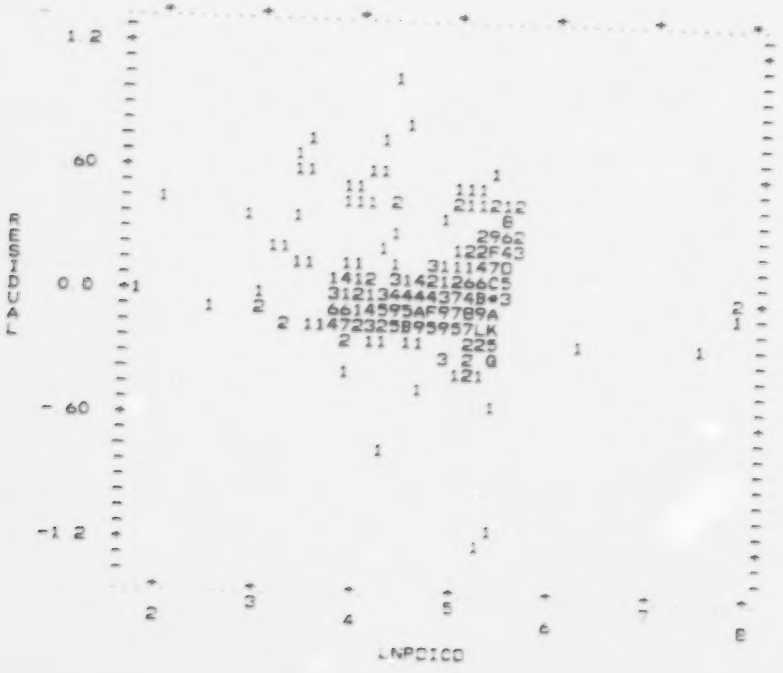


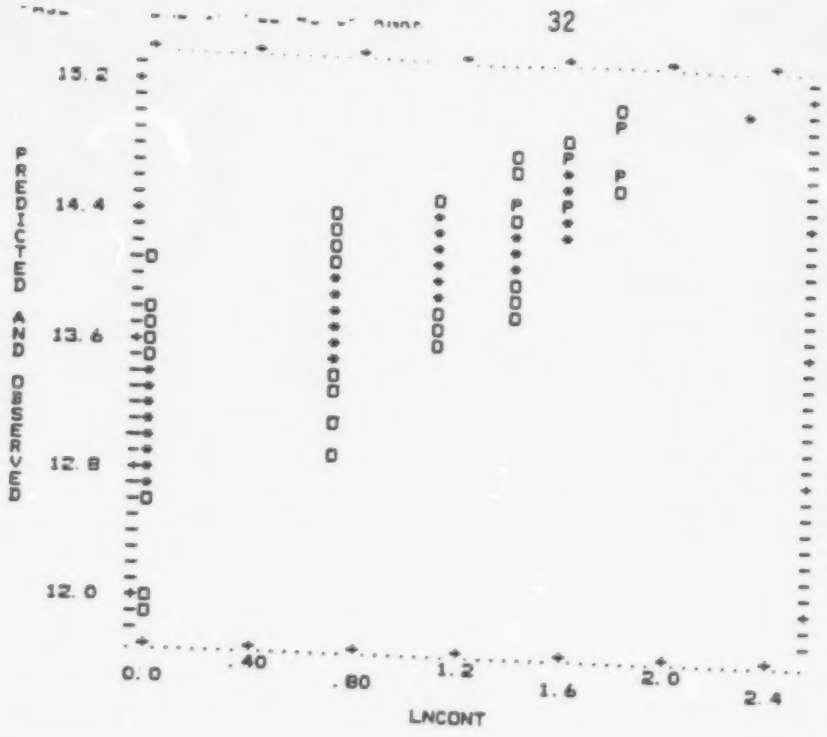
RNEP / ECL / CH / Tate = 1
Rega = 1


PAGE 11 BMDP1R REG B
PREDICTED AND OBSERVED
12000E3
10000E3
8000000
6000000
4000000
2000000
0.00000
-2000000
0.000
1000.
2000.
3000.
4000.
5000.
6000.
7000.
8000.
9000.
10000
PDID

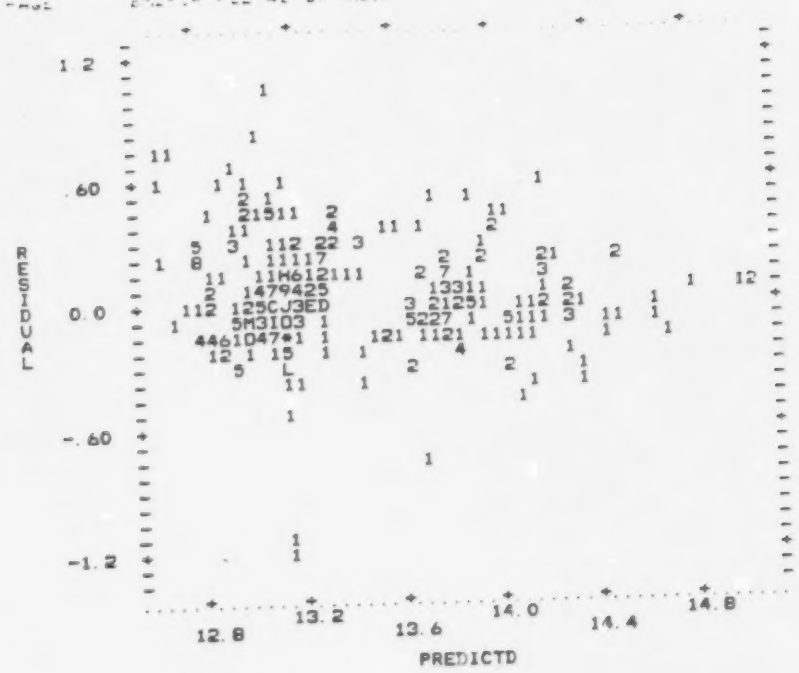


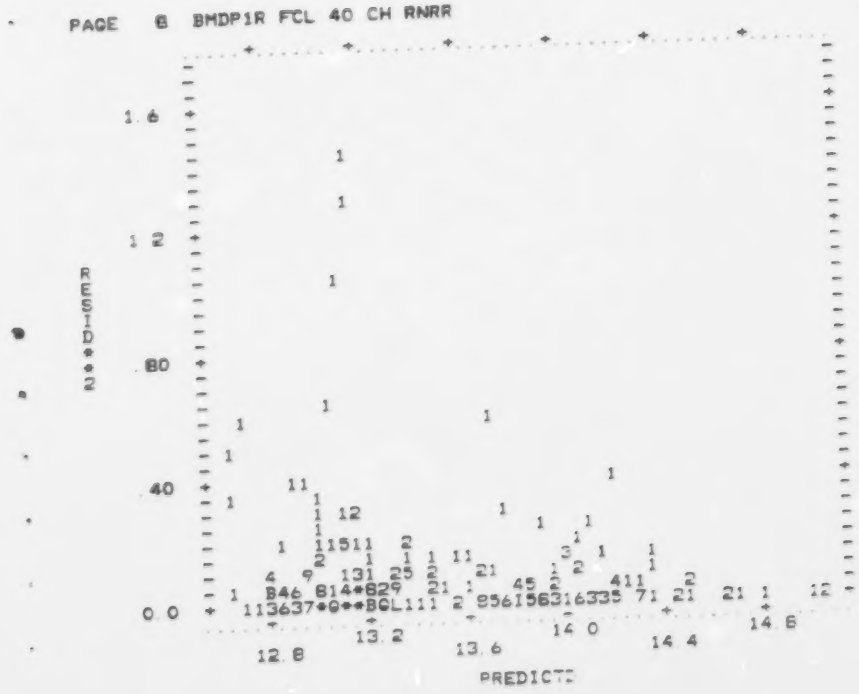
PAGE 12 BMDP1R FCL 40 CH RNRR





PAGE 10 BMDP1R FCL 40 CH RNRR


RESIDUAL
1.2
.60
0.0
-.60
-1.2
12.8
13.2
13.6
14.0
14.4
14.8
PREDICTD


PAGE 8 BMDP1R FCL 40 CH RNRR
RESID
1.6
1.2
.80
.40
0.0
12.8
13.2
13.6
14.0
14.4
14.8
PREDICTD

FCL RNRR 40'

```
                2.6     3 ***
                3.0     0
                3.2     1 *
                3.4     1 *
                3.6     1 *
                3.8     0
GREATER THAN    4       1 *
```

25.81 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 3283, CASE LABEL =

-5.56 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 6116, CASE LABEL =

.06 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 3525, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 1.95 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	10.173841	10.084580	.0088	.4463
12 LNCONT	.969085	.972138	-.0032	-.1422
11 LNKMDIS	.394077	.405126	-.0280	-.3587
14 LNPOICO	.070924	.074031	-.0438	-.2184

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM		
COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.476575E-01	.476575E-01	-.228299E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11	LNPOICO 14
LNCONT	12	1.000		
LNKMDIS	11	.148	1.000	
LNPOICO	14	.021	-.199	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	.0000
AVERAGE RESIDUAL	.04765753
RESIDUAL MEAN SQUARE	.0003
AVERAGE DELETED RESIDUAL	.04808301
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.4597
SERIAL CORRELATION	1.0726
DURBIN-WATSON STATISTIC	

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

```
  LESS THAN -4     2 :**
            -3.6   1 *
            -3.4   0
            -3.2   0
            -3.0   0
            -2.8   0
            -2.6   1 *
            -2.4   1 *
            -2.2   0
            -2.0   0
            -1.8   4 ****
            -1.6   7 *******
            -1.4  26 **************************
            -1.2  11 ***********
            -1.0  17 *****************
             -.8  60 ************************************************************
             -.6  73 *************************************************************************
             -.4  59 ***********************************************************
             -.2  45 *********************************************
              .0  62 ***************************************************************
              .2  36 ************************************
              .4  13 *************
              .6  34 ***********************************
              .8  21 *********************
             1.0  16 ****************
             1.2  12 ************
             1.4  10 **********
             1.6   8 ********
             1.8  12 ************
             2.0   5 *****
```

PAGE 7 BMDP9R FCL RNRR 40 PIEDS

STATISTICS FOR 'BEST' SUBSET

MALLOWS' CP	4.00
SQUARED MULTIPLE CORRELATION	.79030
MULTIPLE CORRELATION	.88899
ADJUSTED SQUARED MULT. CORR.	.78918
RESIDUAL MEAN SQUARE	.047658
STANDARD ERROR OF EST.	.218306
F-STATISTIC	700.99
NUMERATOR DEGREES OF FREEDOM	3
DENOMINATOR DEGREES OF FREEDOM	558
SIGNIFICANCE (TAIL PROB.)	.0000

NOTE THAT THE ABOVE F-STATISTIC AND ASSOCIATED SIGNIFICANCE TEND TO BE LIBERAL WHENEVER A SUBSET OF VARIABLES IS SELECTED BY THE CP OR ADJUSTED R-SQUARED CRITERIA.

NO.	VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ
	INTERCEPT	10.1738	.199998	21.398	50.87	.000		
12	LNCONT	.969085	.0214792	.886	45.12	.000	.975340	.76497
11	LNKMDIS	.394077	.0308011	.256	12.79	.000	.937191	.06152
14	LNPOICO	.0709242	.0142258	.099	4.99	.000	.957873	.00934

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

PAGE 8 BMDP9R FCL RNRR 40 PIEOS

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11	LNPOICO 14
LNCONT	12	1.000		
LNKMDIS	11	.148	1.000	
LNPOICO	14	.021	-.199	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	.0000
AVERAGE RESIDUAL	.04765753
RESIDUAL MEAN SQUARE	.0003

PAGE 4 BMDP9R FCL RNRR 40 PIEDS

NUMBER OF CASES READ 11248
CASES WITH USE SET TO ZERO 10686
REMAINING NUMBER OF CASES 562

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	.22907	.43450	1.896807	.00000	2.30259	-.53
11 LNKMDIS	6.51529	.30910	.047443	5.56452	7.18007	-3.08
14 LNPOICO	4.91720	.66199	.134628	1.79176	7.86327	-4.72
13 LNDEP	13.31211	.47545	.035716	11.94860	15.13534	-2.07

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

PAGE 5 BMDP9R FCL RNRR 40 PIEDS

CORRELATIONS

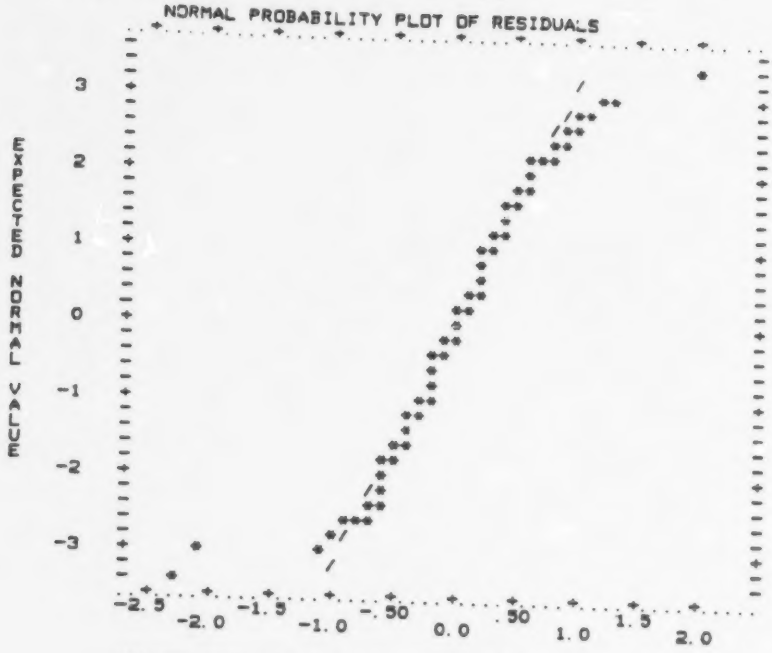
		LNCONT 12	LNKMDIS 11	LNPOICO 14	LNDEP 13
LNCONT	12	1.000			
LNKMDIS	11	-.156	1.000		
LNPOICO	14	-.052	.204	1.000	
LNDEP	13	.841	.139	.105	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

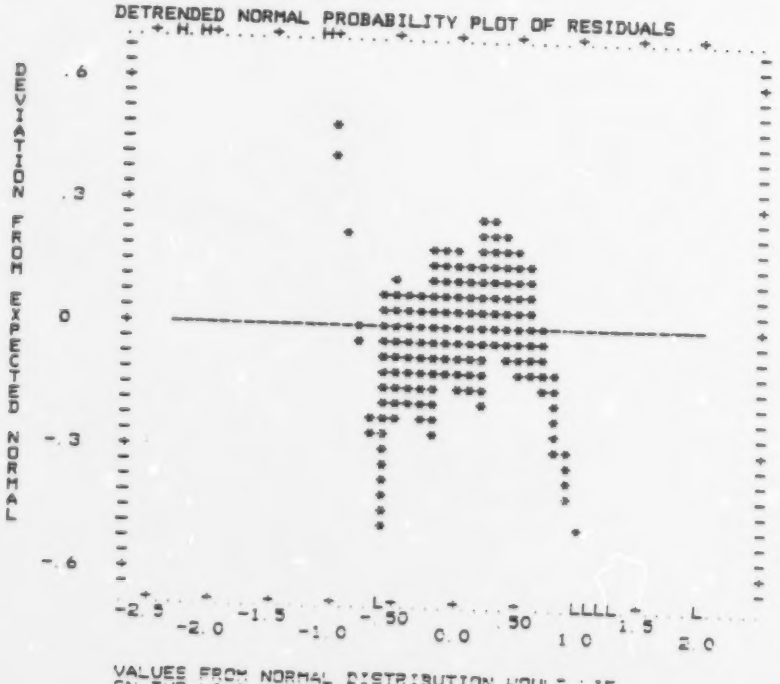
```
12 LNCONT    *
13 LNDEP     **
11 LNKMDIS   . .*
14 LNPOICO     .*
```

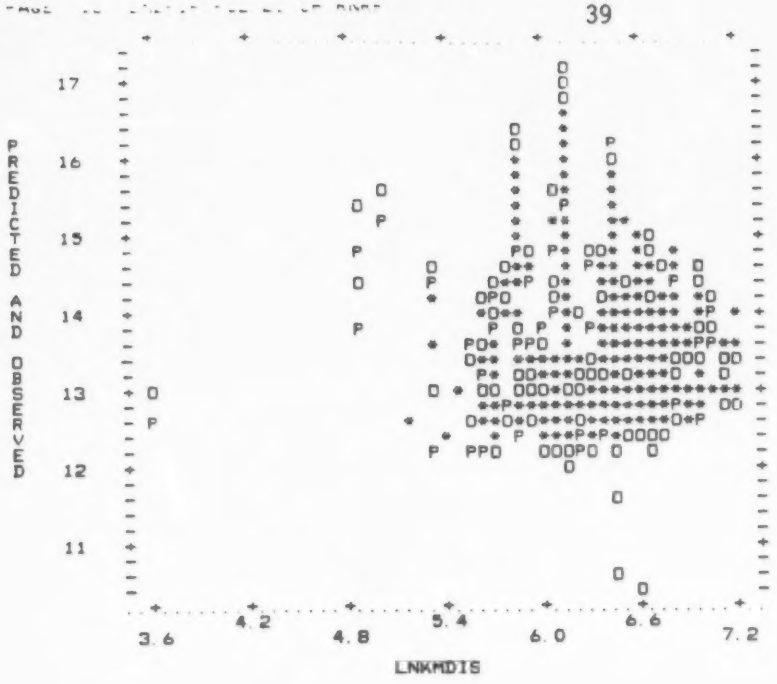
THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

	LESS THAN OR EQUAL TO		
.	.105	TO AND INCLUDING	.105
-	.210	TO AND INCLUDING	.210
+	.315	TO AND INCLUDING	.315
X	.420	TO AND INCLUDING	.420
N	.525	TO AND INCLUDING	.525
C	.630	TO AND INCLUDING	.630
		GREATER THAN	.736

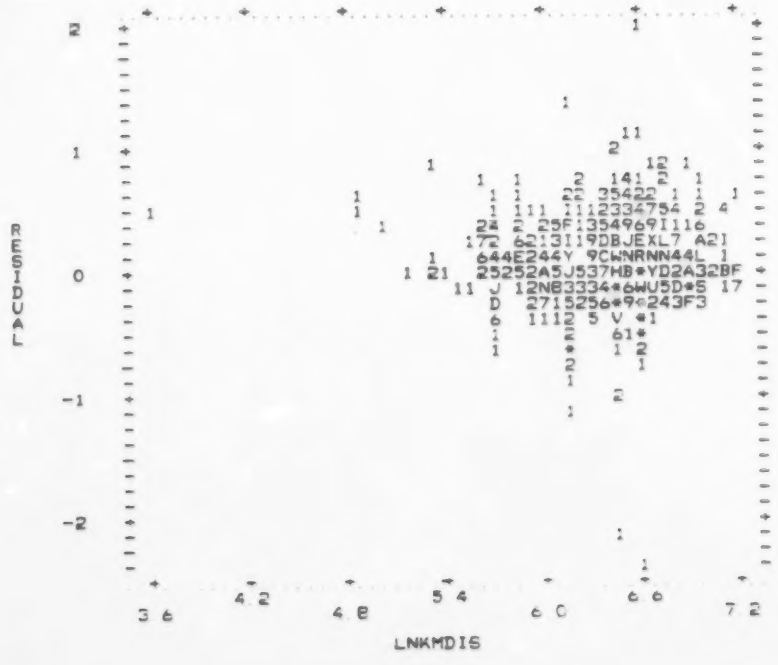


PAGE 21 BMDP1R FCL 20 CH RNRR

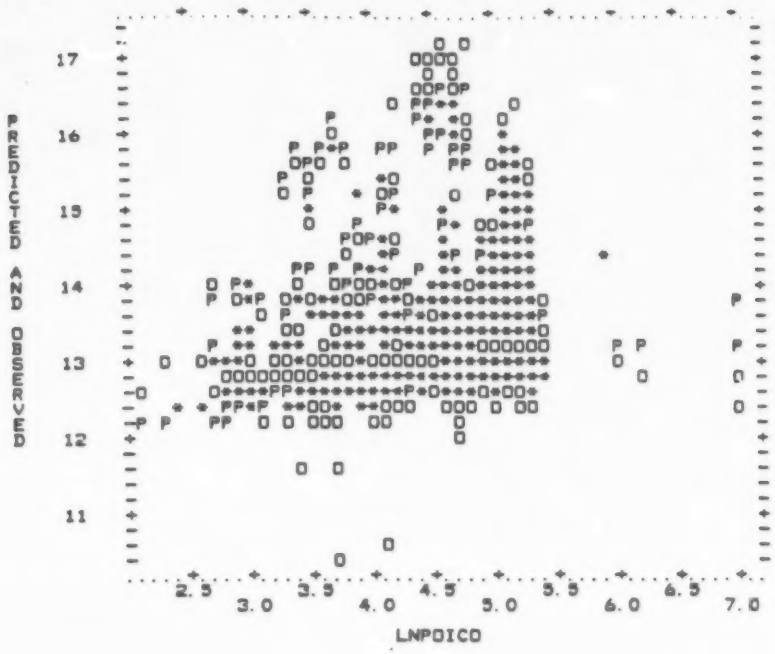




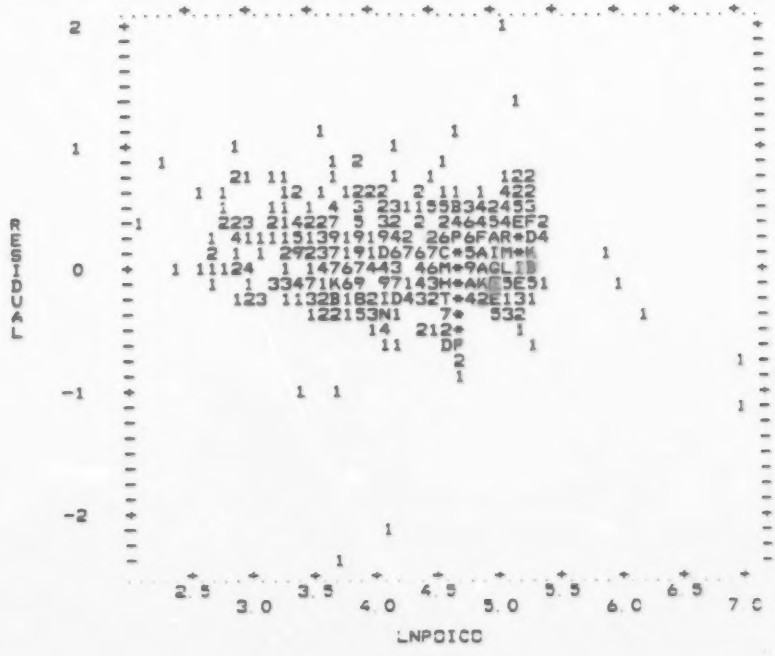
PAGE 14 BMDP1R FCL 20 CH RNRR

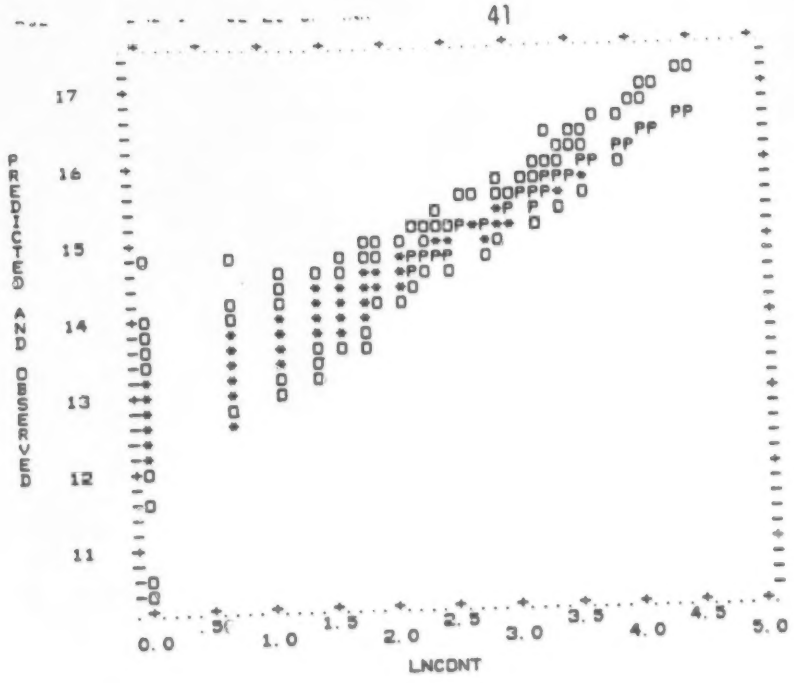


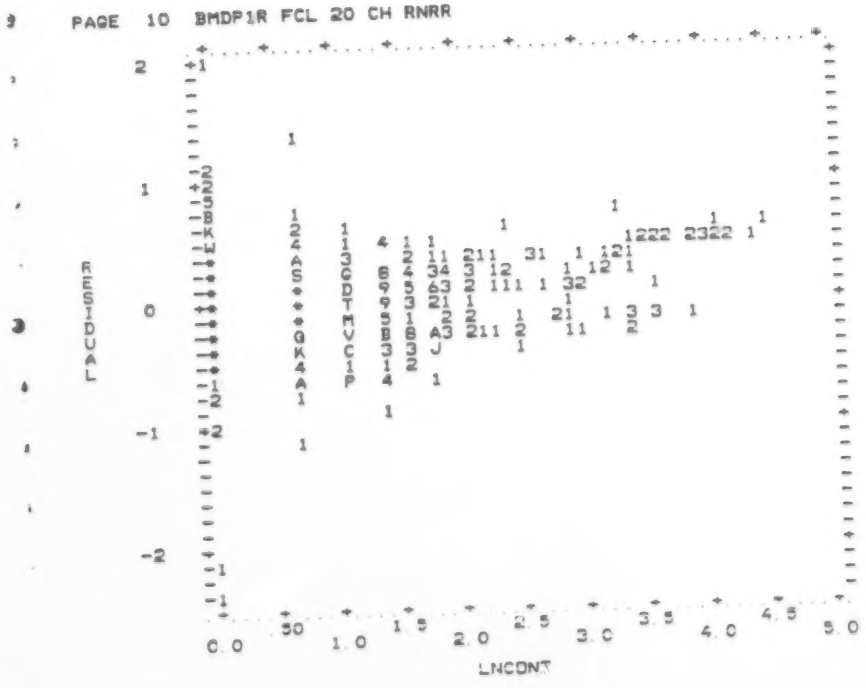
PAGE 11 BMDP1R FCL 20 CH RNRR

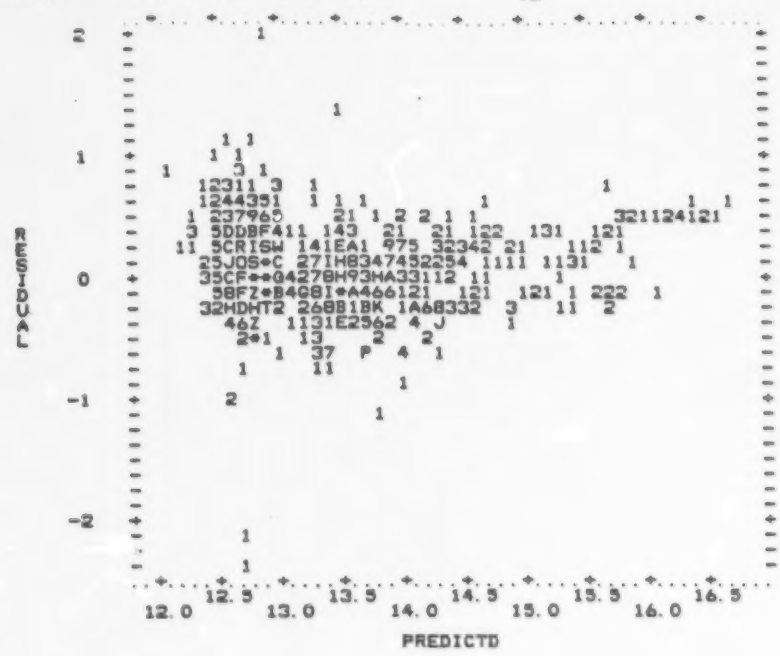


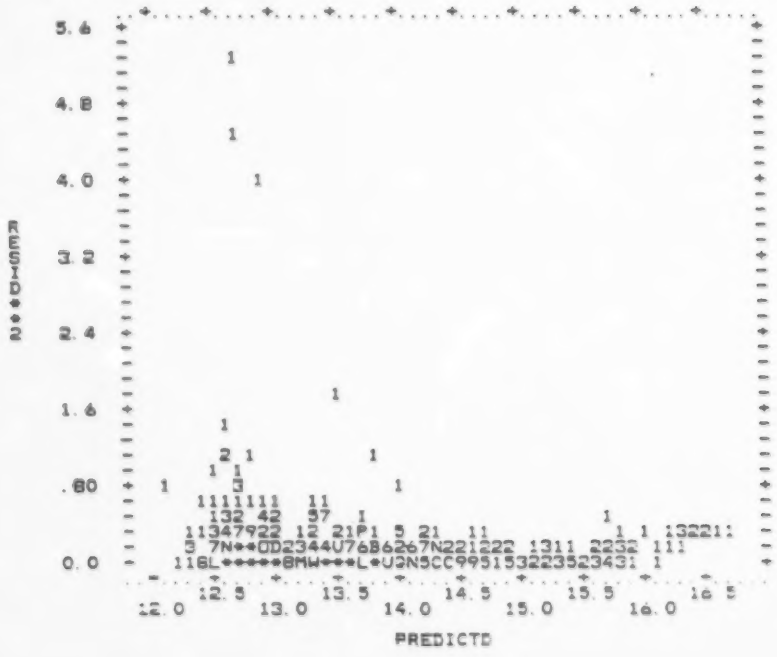
PAGE 12 BMDP1R FCL 20 CH RNRR

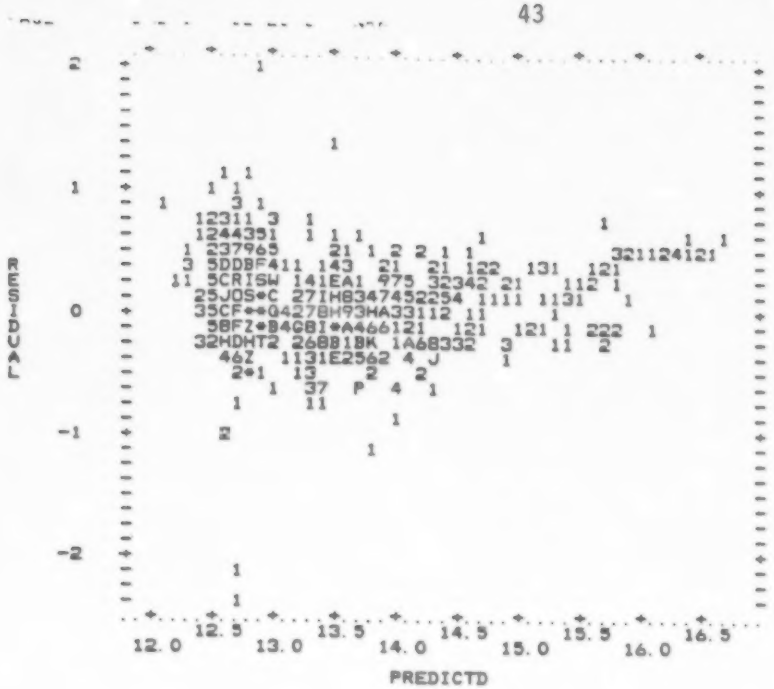



PREDICTED AND OBSERVED
17
16
15
14
13
12
11
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
LNCONT

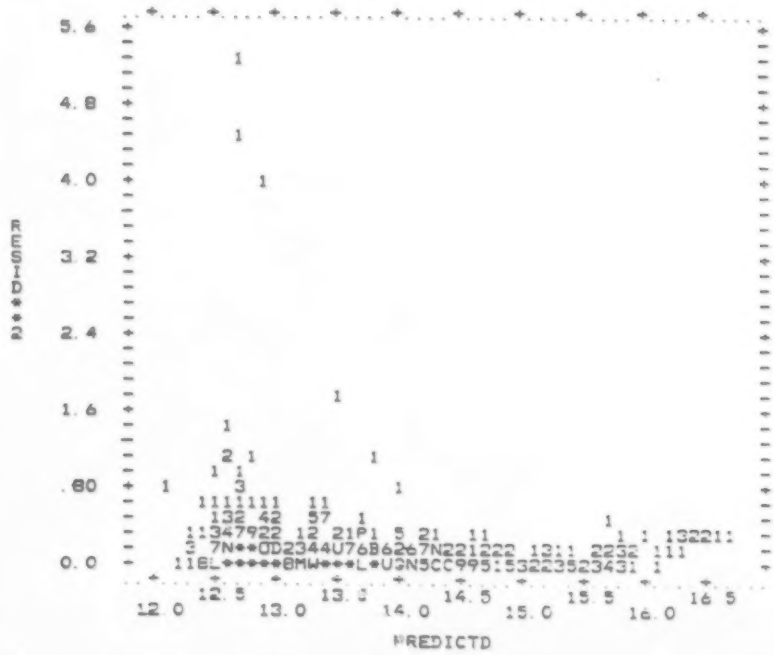

RESIDUAL
2
1
0
-1
-2
0.0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
LNCONT


RESIDUAL
2
1
0
-1
-2
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
PREDICTD


RESIDUAL
5.6
4.8
4.0
3.2
2.4
1.6
.80
0.0
12.0
12.5
13.0
13.5
14.0
14.5
15.0
15.5
16.0
PREDICTD



PAGE 6 BMDP1R FCL 20 CH RNRR



```
                2.8     4 ****
                3.0     2 **
                3.2     0
                3.4     1 *
                3.6     1 *
                3.8     0
GREATER THAN 4          2 **
```

70.92 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 5083, CASE LABEL =

-7.51 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 6127, CASE LABEL =

.04 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 9303, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 1.06 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	10.498193	10.441499	.0054	.3590
12 LNCONT	.889938	.890985	-.0012	-.1053
11 LNKMDIS	.224668	.230539	-.0261	-.2664
14 LNPOICO	.184282	.188130	-.0209	-.2894

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM

COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.952601E-01	.952601E-01	.303021E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

```
LNCONT     12     1.000
LNKMDIS    11      .262     1.000
LNPOICO    14     -.039      .024     1.000
```

FCL RNRR 20-

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)

AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.09526008
AVERAGE DELETED RESIDUAL	.0002
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.09557247
SERIAL CORRELATION	.5125
DURBIN-WATSON STATISTIC	.9745

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

```
  LESS THAN -4     2 :**
          -3.6     1 *
          -3.4     2 **
          -3.2     0
          -3.0     0
          -2.8     1 *
          -2.6     2 **
          -2.4     1 *
          -2.2    11 ***********
          -2.0    30 ******************************
          -1.8     4 ****
          -1.6    56 ********************************************************
          -1.4    29 *****************************
          -1.2    77 *****************************************************************************
          -1.0    79 *******************************************************************************
          -.8    100 ****************************************************************************************************
          -.6    147 ****************************************************************************************************
          -.4    135 ****************************************************************************************************
          -.2    117 ****************************************************************************************************
           .0    103 ****************************************************************************************************
           .2    118 ****************************************************************************************************
           .4    129 ****************************************************************************************************
           .6    124 ****************************************************************************************************
           .8     77 *****************************************************************************
          1.0     46 ********************************************
          1.2     49 *************************************************
          1.4     34 **********************************
          1.6     27 ***************************
          1.8     18 ******************
          2.0      9 *********
          2.2      7 *******
          2.4      5 *****
          2.6      2 **
          2.8      4 ****
          3.0      2 **
          3.2      0
          3.4      1 *
          3.6      1 *
          3.8      0
GREATER THAN 4     2 **
```

70 92 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH ... 5083. CASE LABEL =

PAGE 7 BMDP9R FCL RNRR 20 PIEDS

STATISTICS FOR 'BEST' SUBSET

MALLOWS' CP	4.00
SQUARED MULTIPLE CORRELATION	.84613
MULTIPLE CORRELATION	.91986
ADJUSTED SQUARED MULT. CORR.	.84584
RESIDUAL MEAN SQUARE	.095260
STANDARD ERROR OF EST.	.308642
F-STATISTIC	2837.59
NUMERATOR DEGREES OF FREEDOM	3
DENOMINATOR DEGREES OF FREEDOM	1548
SIGNIFICANCE (TAIL PROB.)	.0000

NOTE THAT THE ABOVE F-STATISTIC AND ASSOCIATED SIGNIFICANCE TEND TO BE LIBERAL WHENEVER A SUBSET OF VARIABLES IS SELECTED BY THE CP OR ADJUSTED R-SQUARED CRITERIA.

NO.	VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ
	INTERCEPT	10.4982	.157927	13.355	66.47	.000		
12	LNCONT	.889938	.00994772	.925	89.46	.000	.929185	.79550
11	LNKMDIS	.224668	.0220433	.105	10.19	.000	.930087	.01033
14	LNPOICO	.184282	.0132976	.138	13.86	.000	.997158	.01909

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

PAGE 8 BMDP9R FCL RNRR 20 PIEDS

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11	LNPOICO 14
LNCONT	12	1.000		
LNKMDIS	11	-.262	1.000	
LNPOICO	14	-.039	.024	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.09526008
AVERAGE DELETED RESIDUAL	.0002

NUMBER OF CASES READ. 11248
CASES WITH USE SET TO ZERO 9696
REMAINING NUMBER OF CASES 1552

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	.54760	.81729	1.492478	.00000	4.46591	-.67
11 LNKMDIS	6.47854	.36865	.056903	3.58352	7.18007	-7.05
14 LNPOICO	4.54652	.59020	.129813	2.07944	6.95655	-4.10
13 LNDEP	13.27889	.78608	.059198	10.35201	17.29158	-3.72

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

CORRELATIONS

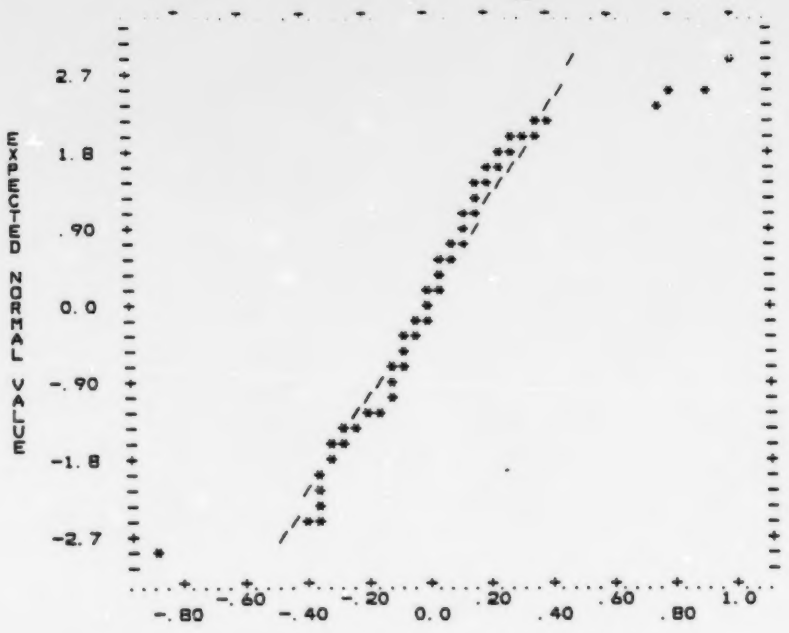
		LNCONT 12	LNKMDIS 11	LNPOICO 14	LNDEP 13
LNCONT	12	1.000			
LNKMDIS	11	-.263	1.000		
LNPOICO	14	.047	-.036	1.000	
LNDEP	13	.904	-.143	.179	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

```
12 LNCONT    *
13 LNDEP     **
11 LNKMDIS   -.*
14 LNPOICO    . *
```

THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

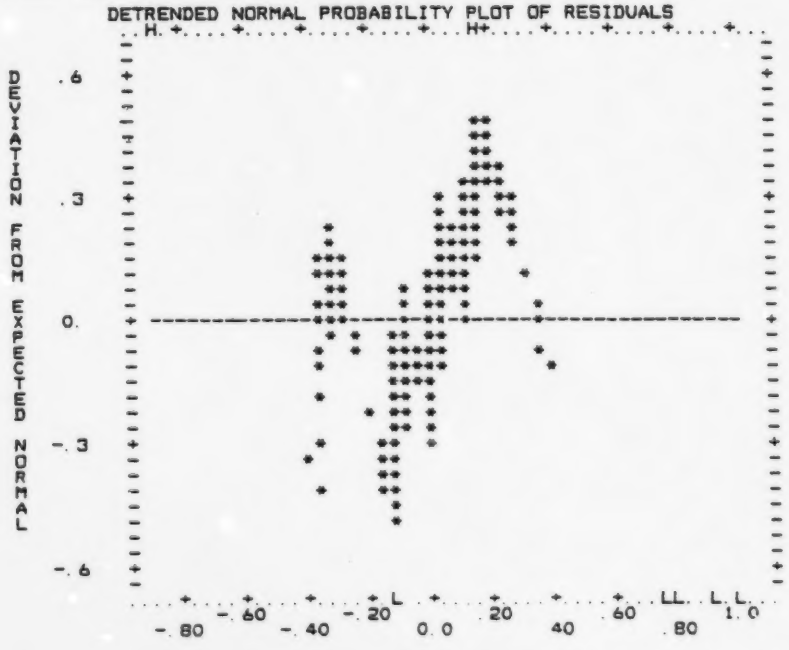
	LESS THAN OR EQUAL TO		
	.113	TO AND INCLUDING	.226
-	.226	TO AND INCLUDING	.339
+	.339	TO AND INCLUDING	.452
X	.452	TO AND INCLUDING	.565
N	.565	TO AND INCLUDING	.678
O	.678	TO AND INCLUDING	.791
*		GREATER THAN	.791



VALUES FROM NORMAL DISTRIBUTION WOULD LIE
ON THE LINE INDICATED BY THE SYMBOL /

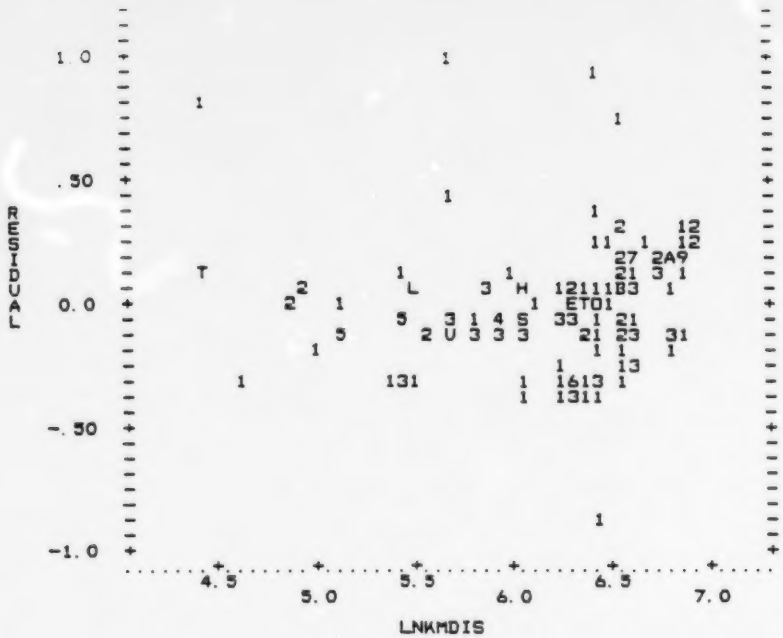
PAGE 18 BMDP1R VIDE 40 FM RNEP

DETRENDED NORMAL PROBABILITY PLOT OF RESIDUALS

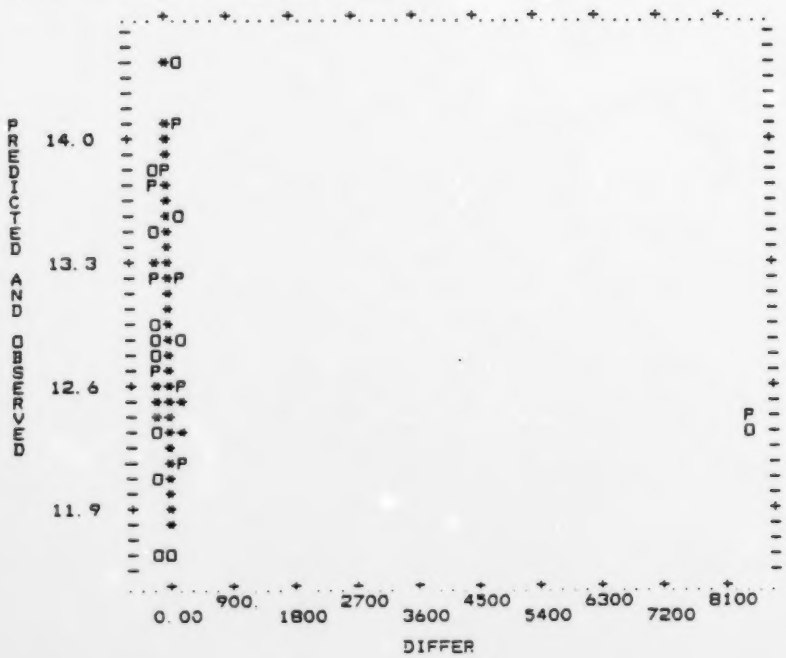


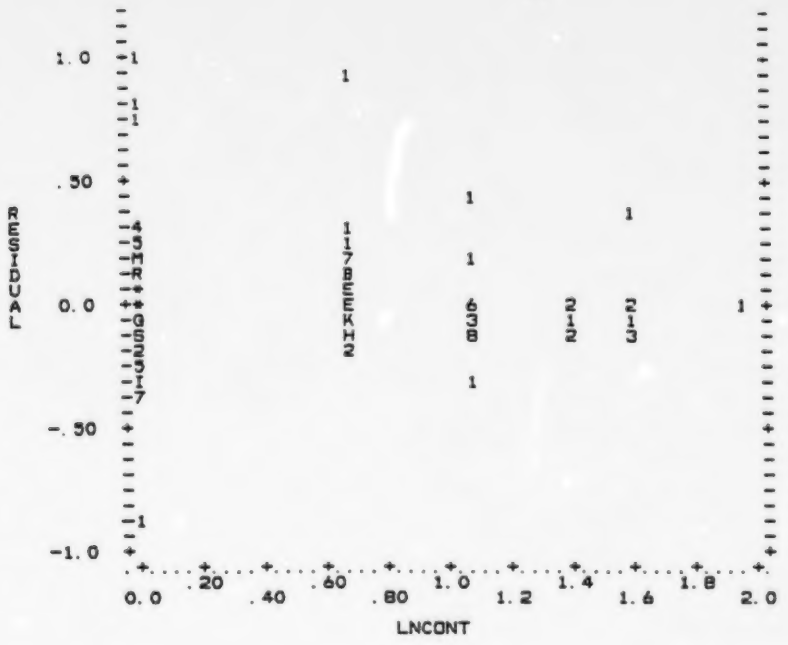
VALUES FROM NORMAL DISTRIBUTION WOULD LIE
ON THE LINE INDICATED BY THE SYMBOL -

NUMBER OF INTEGER WORDS OF STORAGE USED IN PREPARING PROBLEM

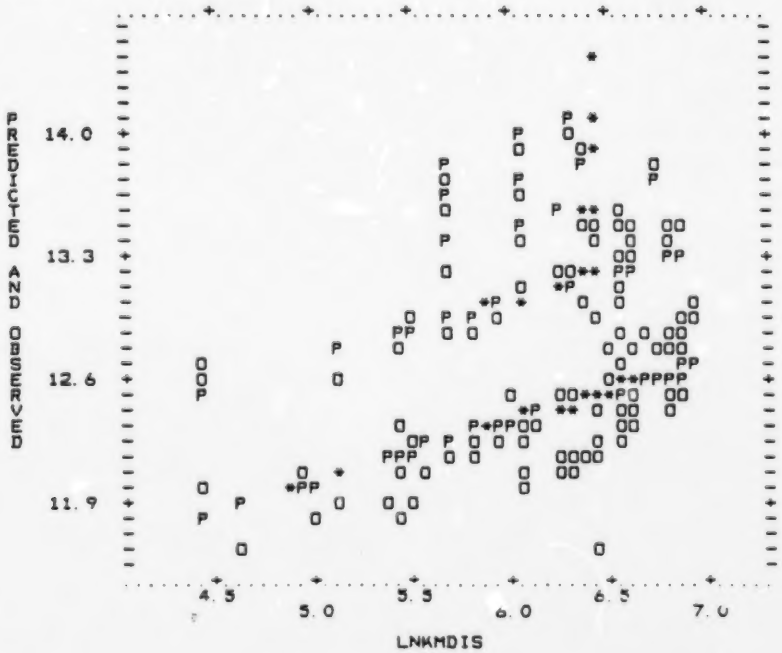


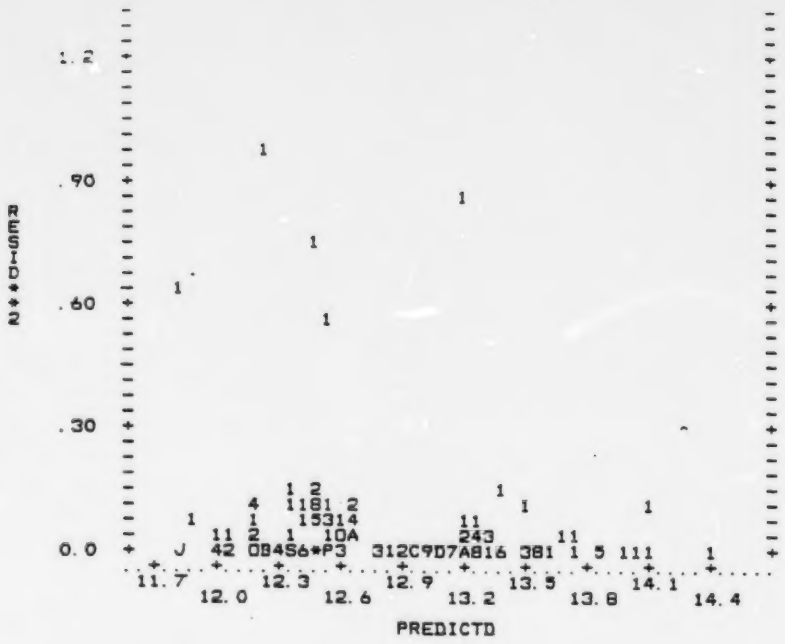
PAGE 13 BMDP1R VIDE 40 FM RNEP



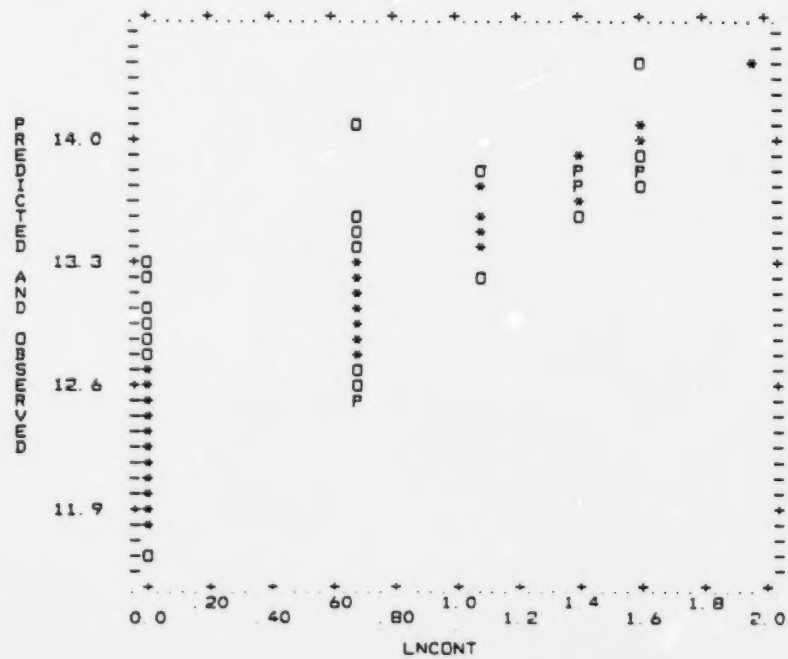


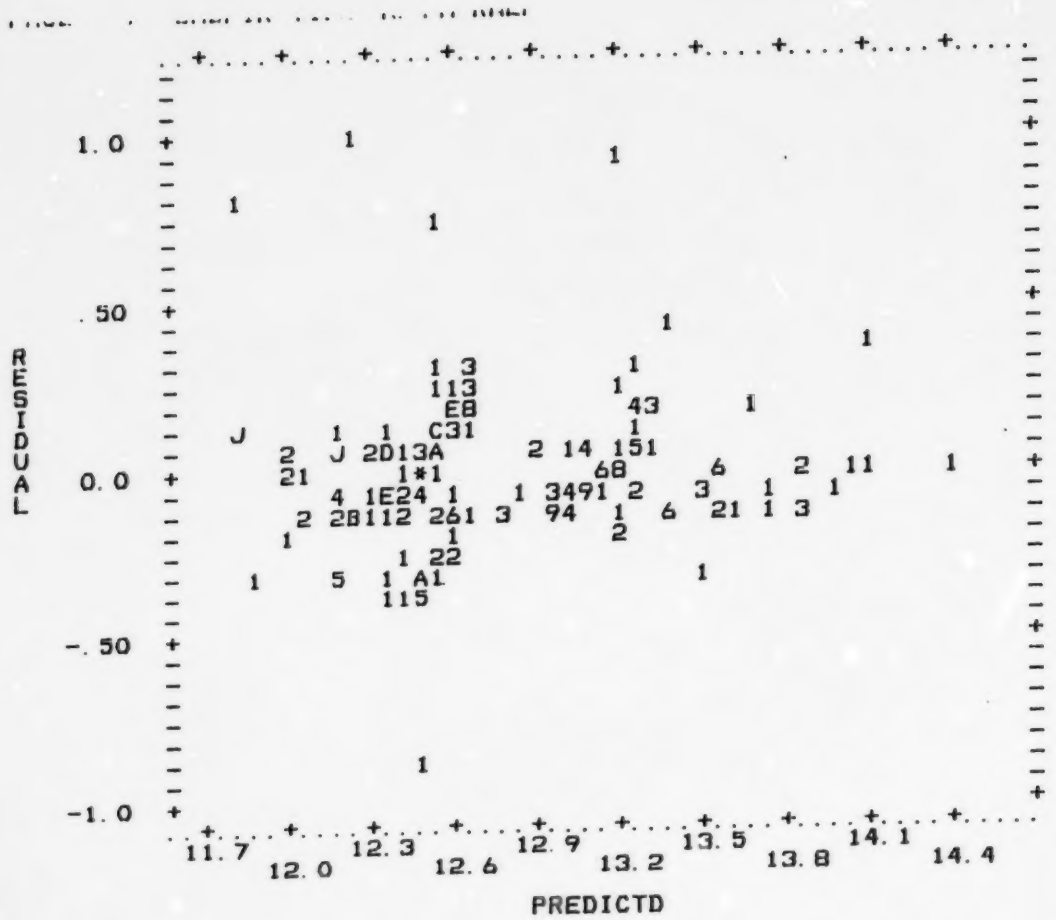
PAGE 11 BMDP1R VIDE 40 FM RNEP





PAGE 9 BMDP1R VIDE 40 FM RNEP




RESIDUAL
1.0
.50
0.0
-.50
-1.0
11.7
12.0
12.3
12.6
12.9
13.2
13.5
13.8
14.1
14.4
PREDICTD

15.09 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 2100, CASE LABEL =

5.65 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 3915, CASE LABEL =

.15 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 4251, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 14.33 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	10.354317	10.299234	.0053	.6465
12 LNCONT	.988330	.991885	-.0036	-.1696
11 LNKMDIS	.334348	.342934	-.0257	-.6139

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM		
COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.302910E-01	.302910E-01	-.234572E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11
LNCONT	12	1.000	
LNKMDIS	11	.014	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.03029102
AVERAGE DELETED RESIDUAL	.0002
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.03053042
SERIAL CORRELATION	.2994
DURBIN-WATSON STATISTIC	1.3930

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

```
  LESS THAN -4     1 :*
            -2.4   1 *
            -2.2   5 *****
            -2.0  11 ***********
            -1.8   9 *********
            -1.6   1 *
            -1.4   4 ****
            -1.2   2 **
            -1.0  10 **********
             -.8  43 ******************************************
             -.6  37 ************************************
             -.4  20 ********************
             -.2   4 ****
              .0  85 *************************************************************************************
              .2  41 ****************************************
              .4   8 ********
              .6  32 ********************************
              .8  14 **************
             1.0  18 ******************
             1.2   4 ****
             1.4   5 *****
             1.6   2 **
             1.8   3 ***
             2.0   1 *
             2.2   1 *
             2.4   0
             2.6   0
             2.8   0
             3.0   0
```

STATISTICS FOR 'BEST' SUBSET

SQUARED MULTIPLE CORRELATION	.88382
MULTIPLE CORRELATION	.94012
ADJUSTED SQUARED MULT. CORR.	.88318
RESIDUAL MEAN SQUARE	.030291
STANDARD ERROR OF EST.	.174043
F-STATISTIC	1380.68
NUMERATOR DEGREES OF FREEDOM	2
DENOMINATOR DEGREES OF FREEDOM	363
SIGNIFICANCE (TAIL PROB.)	.0000

VARIABLE NO.	NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ
	INTERCEPT	10.3543	.0852057	20.334	121.52	.000		
12	LNCONT	.988330	.0209701	.843	47.13	.000	.999800	.71095
11	LNKMDIS	.334348	.0139868	.428	23.90	.000	.999800	.18289

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

COVARIANCES OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11
LNCONT	12	.0004397	
LNKMDIS	11	.0000041	.0001956

NUMBER OF CASES READ. 11249
CASES WITH USE SET TO ZERO 10883
REMAINING NUMBER OF CASES 366

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	.28173	.43446	1.542134	.00000	1.94591	-.65
11 LNKMDIS	6.03630	.65138	.107910	4.43082	6.95845	-2.46
13 LNDEP	12.65098	.50920	.040250	11.59910	14.44875	-2.07

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

CORRELATIONS

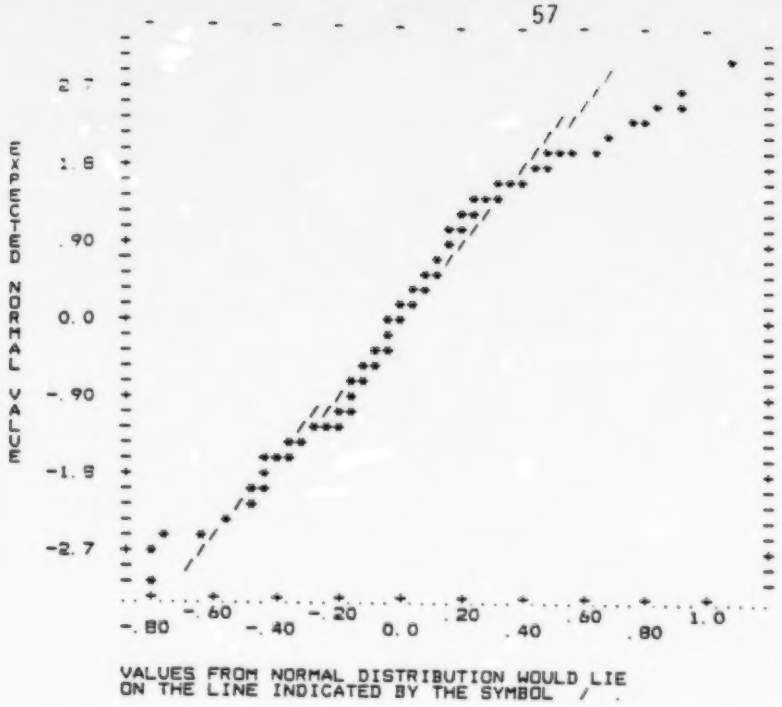
		LNCONT 12	LNKMDIS 11	LNDEP 13
LNCONT	12	1.000		
LNKMDIS	11	-.014	1.000	
LNDEP	13	.837	.416	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

12 LNCONT *
13 LNDEP **
11 LNKMDIS +*

THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

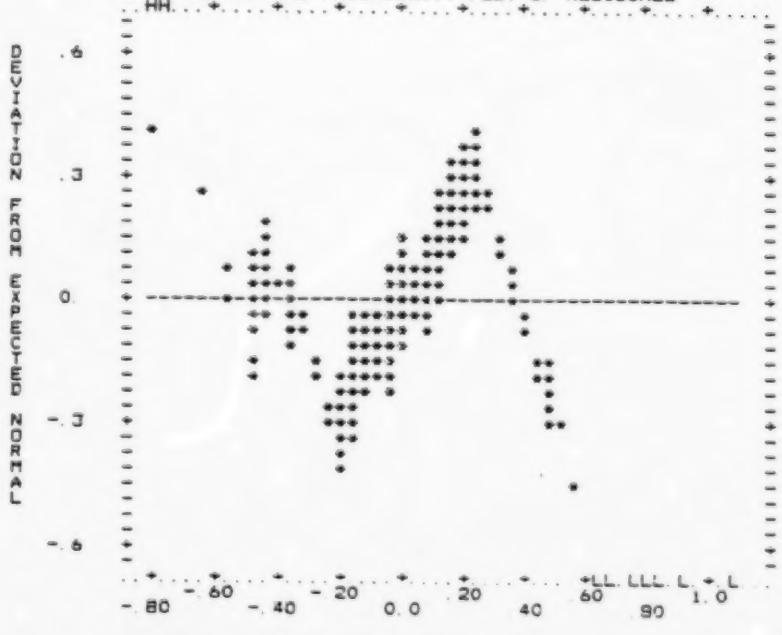
	LESS THAN OR EQUAL TO		.105
.	.105	TO AND INCLUDING	.209
-	.209	TO AND INCLUDING	.314
+	.314	TO AND INCLUDING	.419
X	.419	TO AND INCLUDING	.523
N	.523	TO AND INCLUDING	.628
O	.628	TO AND INCLUDING	.733
*		GREATER THAN	.733



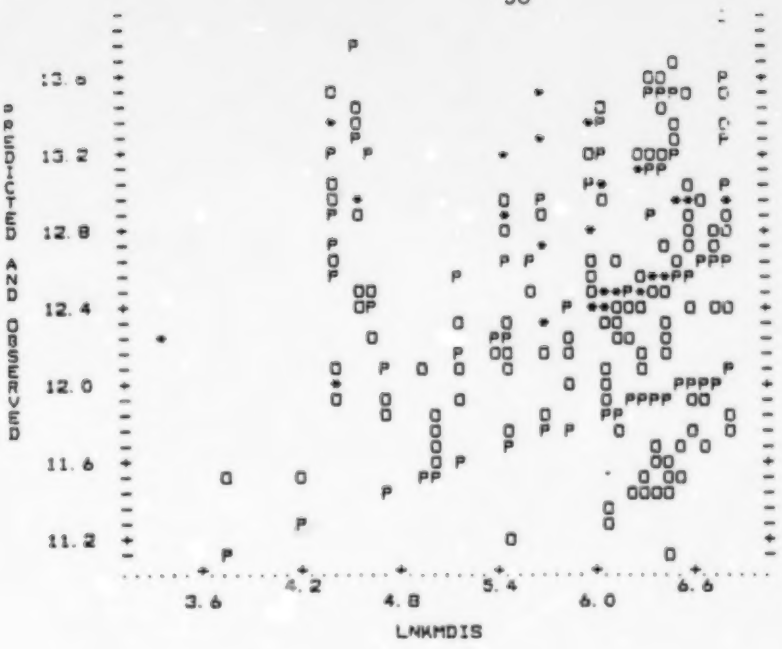
VALUES FROM NORMAL DISTRIBUTION WOULD LIE
ON THE LINE INDICATED BY THE SYMBOL /

PAGE 18 BMDP1R VIDE 20 FM RNEP

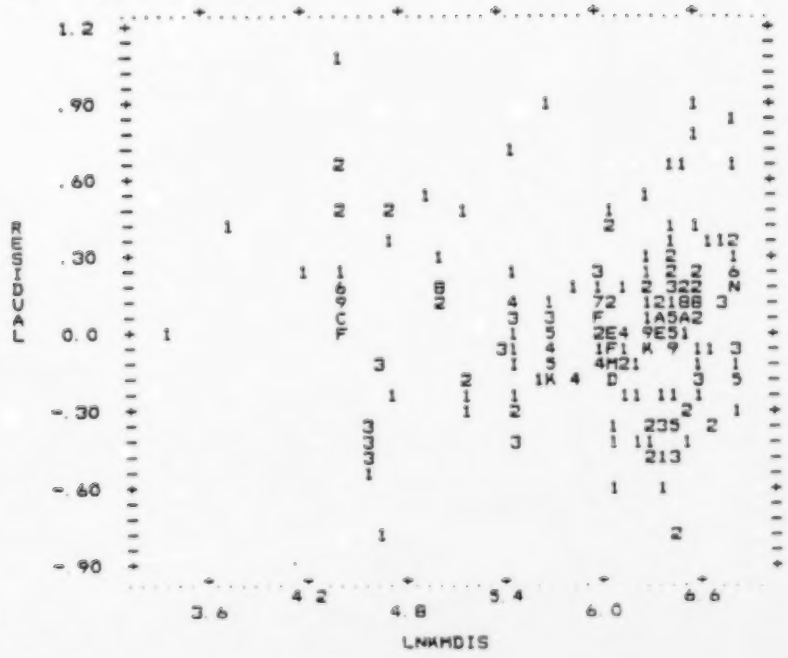
DETRENDED NORMAL PROBABILITY PLOT OF RESIDUALS

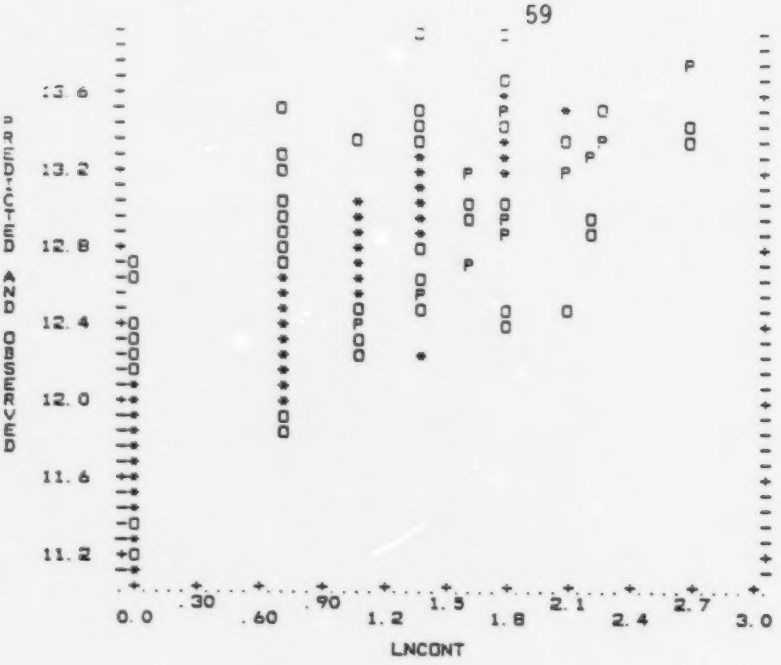


VALUES FROM NORMAL DISTRIBUTION WOULD LIE
ON THE LINE INDICATED BY THE SYMBOL -

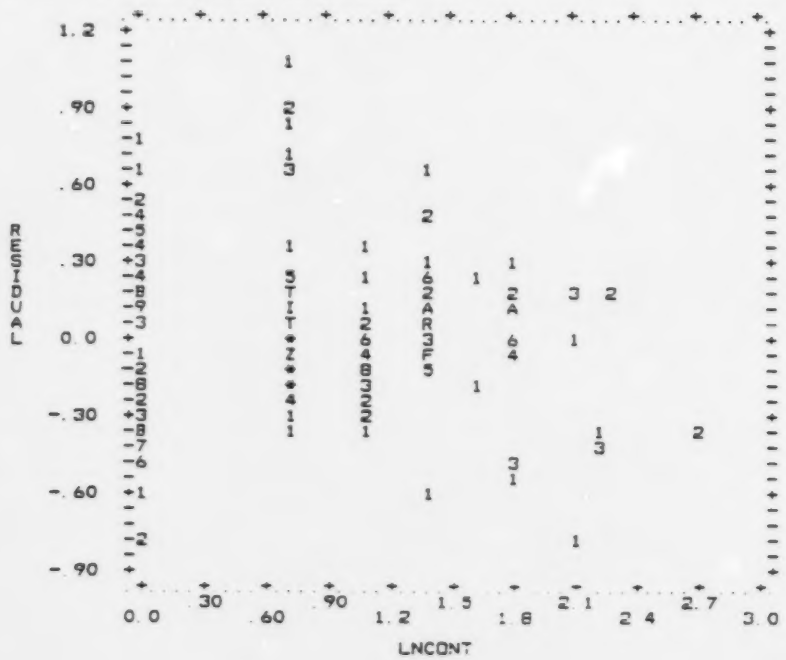


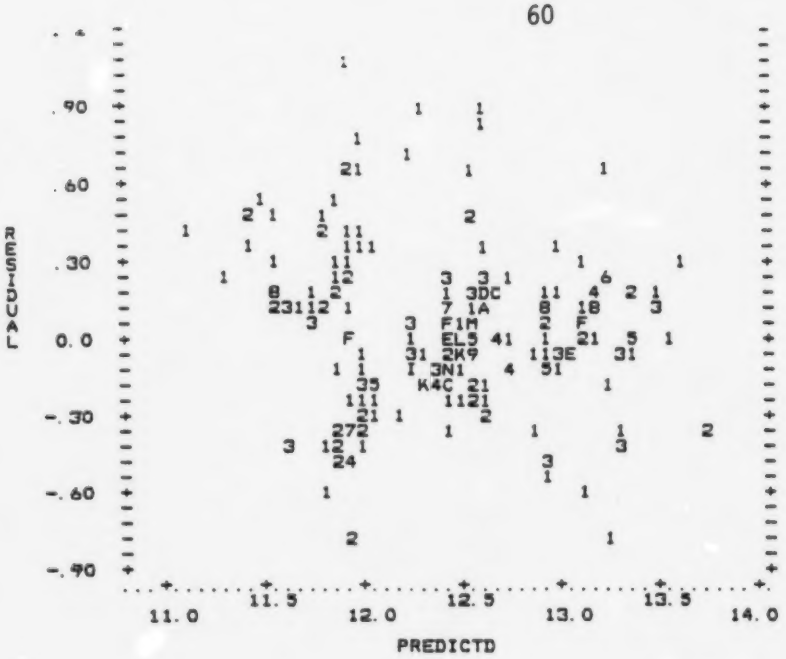
PAGE 12 BMDP1R VIDE 20 FM RNEP



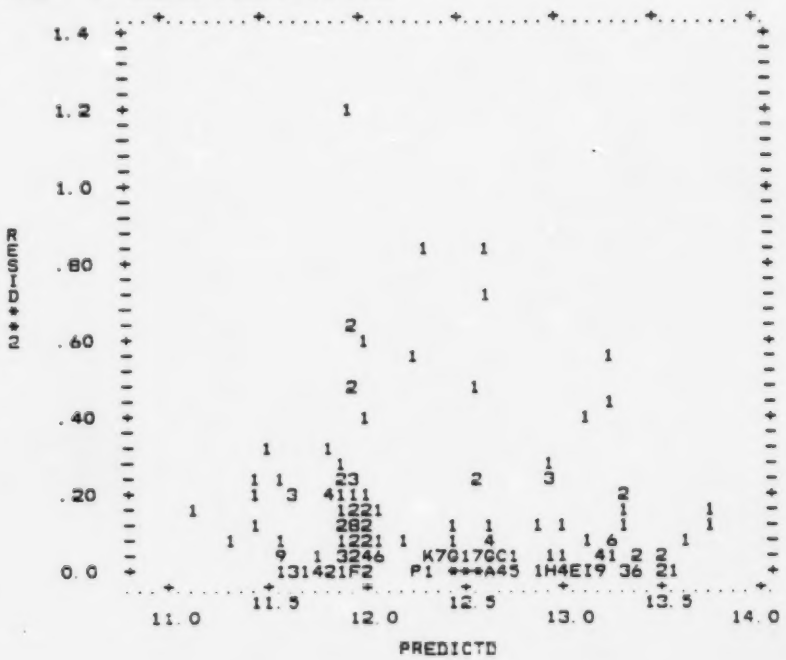


PAGE 10 BMDP1R VIDE 20 FM RNEP





PAGE 8 BMDP1R VIDE 20 FM RNEP



3.8 2 **
GREATER THAN 4 1 *

vide RNEP 20' FM

15.01 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 5342, CASE LABEL =

4.67 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 4259, CASE LABEL =

.09 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 4259, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 8.29 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	10.082730	10.036846	.0046	.4830
12 LNCONT	.868093	.871369	-.0038	-.1621
11 LNKMDIS	.288511	.295435	-.0240	-.4623

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM

COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.547770E-01	.547770E-01	-.328342E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

Vide ANEP 20 FM

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11
LNCONT	12	1.000	
LNKMDIS	11	.265	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.05477702
AVERAGE DELETED RESIDUAL	.0001
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.05533495
SERIAL CORRELATION	.1883
DURBIN-WATSON STATISTIC	1.6230

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

```
-3.4     3 ***
-3.2     0
-3.0     0
-2.8     1 *
-2.6     1 *
-2.4     1 *
-2.2     8 ********
-2.0    10 **********
-1.8     4 ****
-1.6    13 *************
-1.4     4 ****
-1.2     5 *****
-1.0    16 ****************
 -.8    54 ******************************************************
 -.6    40 ****************************************
 -.4    32 ********************************
 -.2    71 ***************************************************************************
  .0    38 ***************************************
  .2    50 **************************************************
  .4    45 *********************************************
  .6    36 ************************************
  .8    20 ********************
 1.0    13 *************
 1.2     4 ****
 1.4     4 ****
 1.6     5 *****
 1.8     4 ****
 2.0     5 *****
 2.2     2 **
 2.4     0
 2.6     1 *
 2.8     4 ****
 3.0     1 *
 3.2     1 *
 3.4     0
```

PAGE 6 BMDP9R FM VIDE RNEP 20 PIEDS

STATISTICS FOR 'BEST' SUBSET

SQUARED MULTIPLE CORRELATION	.79356
MULTIPLE CORRELATION	.89082
ADJUSTED SQUARED MULT. CORR.	.79273
RESIDUAL MEAN SQUARE	.054777
STANDARD ERROR OF EST.	.234045
F-STATISTIC	955.24
NUMERATOR DEGREES OF FREEDOM	2
DENOMINATOR DEGREES OF FREEDOM	497
SIGNIFICANCE (TAIL PROB.)	.0000

NO.	VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOL-ERANCE	CONTRI-BUTION TO R-SQ
	INTERCEPT	10.0827	.0949882	19.613	106.15	.000		
12	LNCONT	.868093	.0202176	.908	42.94	.000	.929803	.76579
11	LNKMDIS	.288511	.0149794	.407	19.26	.000	.929803	.15409

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

PAGE 7 BMDP9R FM VIDE RNEP 20 PIEDS

COVARIANCES OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11
LNCONT	12	.0004088	
LNKMDIS	11	.0000802	.0002244

NUMBER OF CASES READ. 11249
CASES WITH USE SET TO ZERO 10749
REMAINING NUMBER OF CASES 500

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	.79810	.53743	.673387	.00000	2.70805	-1.49
11 LNKMDIS	5.93105	.72537	.122300	3.36730	6.84694	-3.53
13 LNDEP	12.48673	.51408	.041170	11.14186	13.92068	-2.62

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

CORRELATIONS

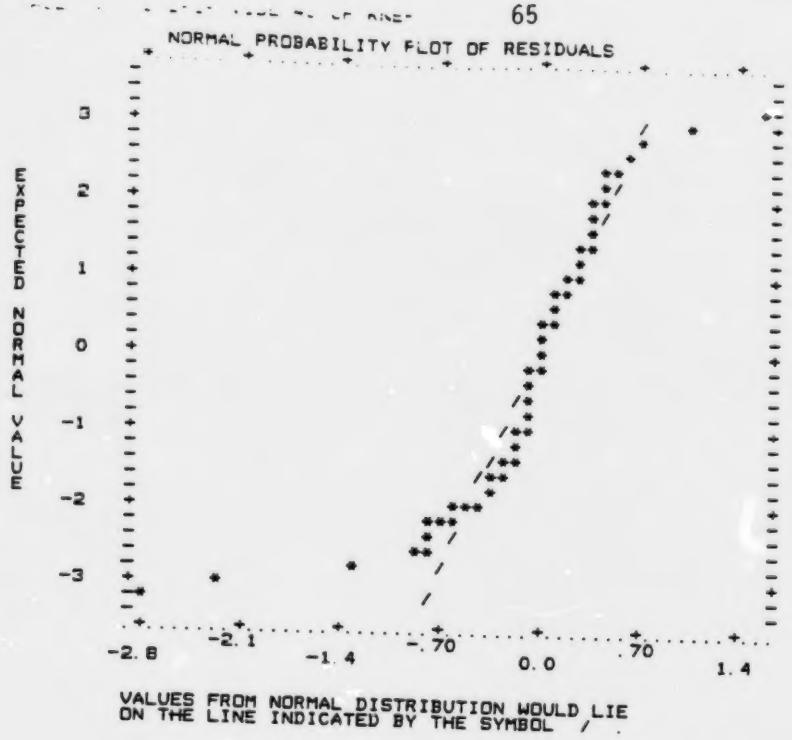
		LNCONT 12	LNKMDIS 11	LNDEP 13
LNCONT	12	1.000		
LNKMDIS	11	-.265	1.000	
LNDEP	13	.800	.167	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

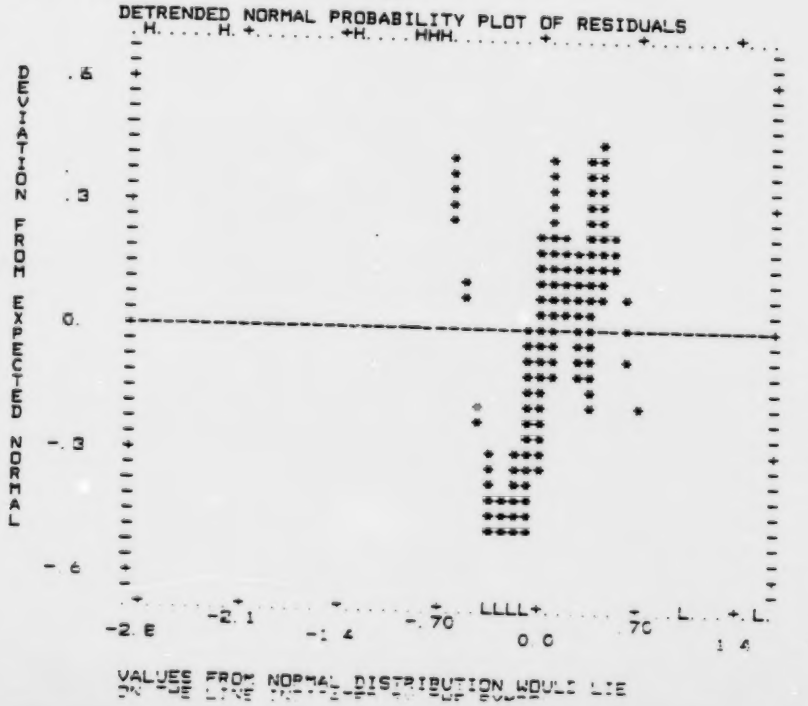
```
  12 LNCONT    *
  13 LNDEP     **
  11 LNKMDIS   -.*
```

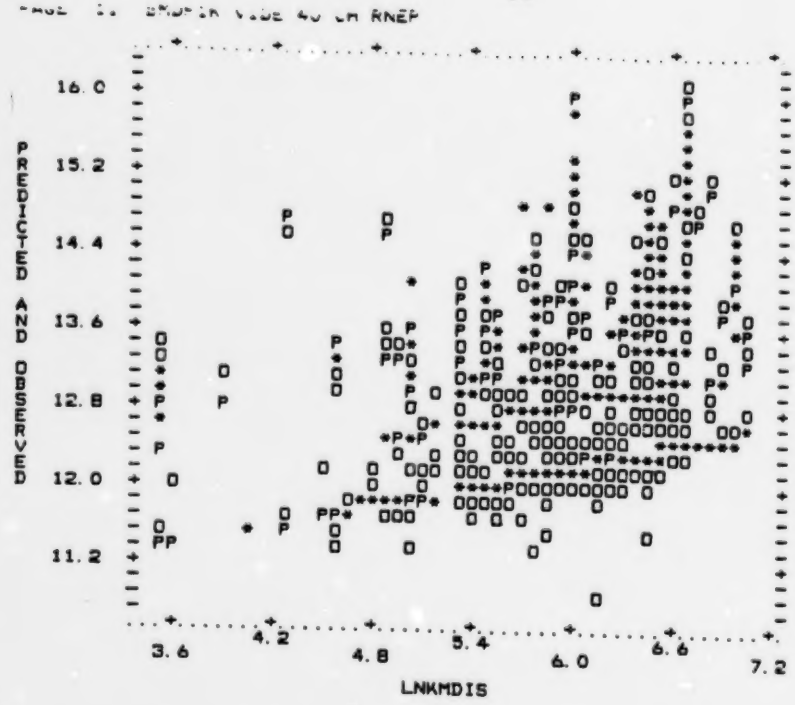
THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

	LESS THAN OR EQUAL TO		.100
.	.100	TO AND INCLUDING	.200
-	.200	TO AND INCLUDING	.300
+	.300	TO AND INCLUDING	.400
X	.400	TO AND INCLUDING	.500
N	.500	TO AND INCLUDING	.600
O	.600	TO AND INCLUDING	.700
*		GREATER THAN	.700

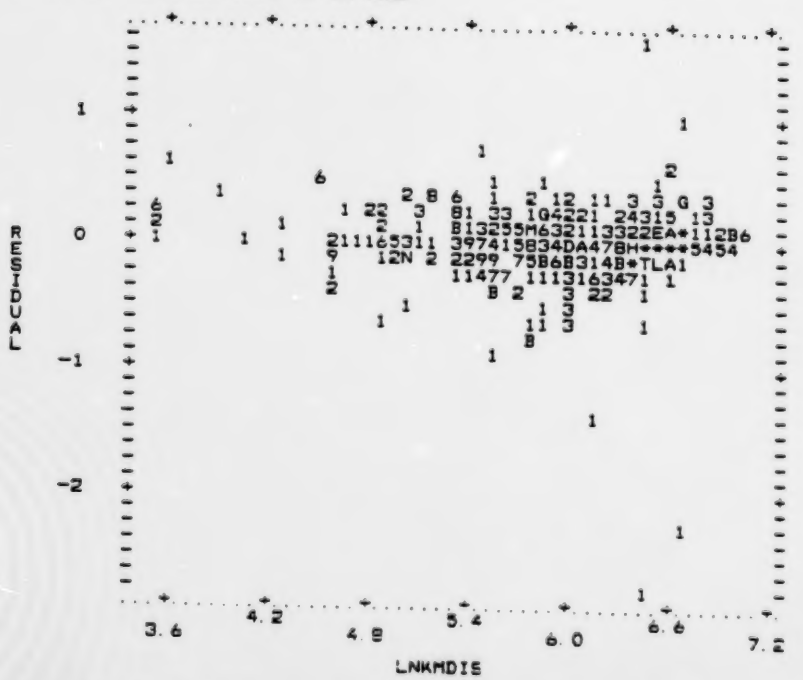


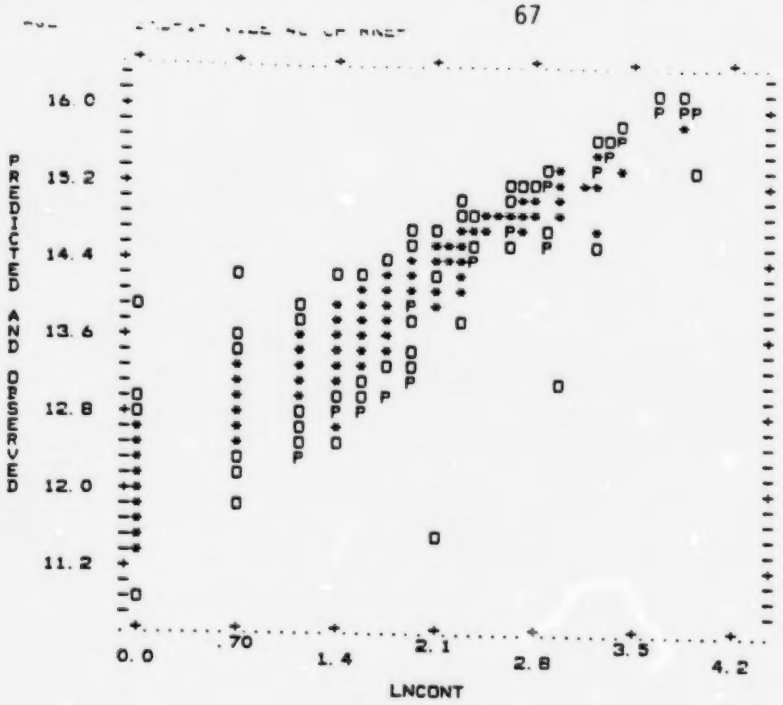
PAGE 18 BMDP1R VIDE 40 CH RNEP



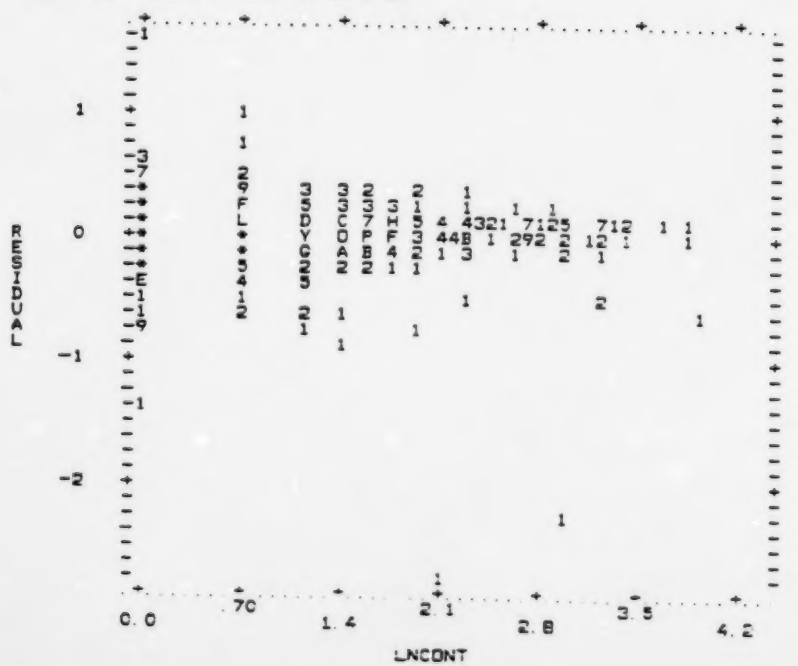


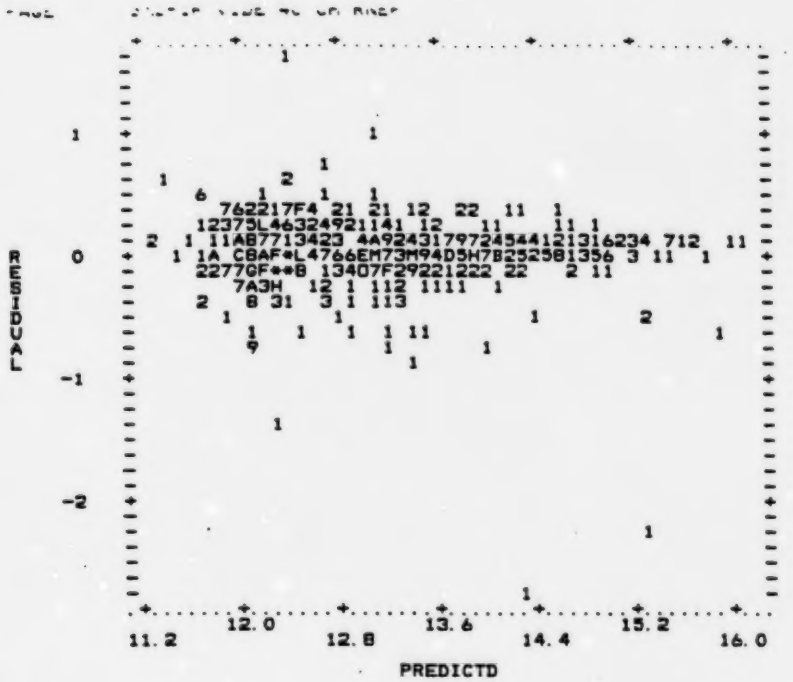
PAGE 12 BMDP1R VIDE 40 CH RNEP



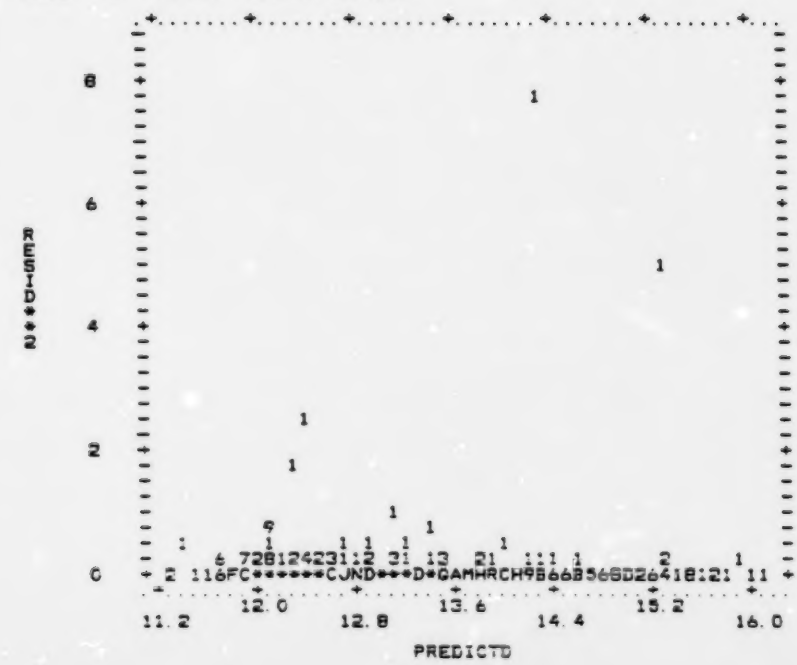


PAGE 10 BMDP1R VIDE 40 CH RNEP





PAGE 8 BMDP1R VIDE 40 CH RNEP



	2.8	0
	3.0	1 *
	3.2	0
	3.4	0
	3.6	0
	3.8	0
GREATER THAN	4	2 **

vide RNEP 40' CH

19.30 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 11227, CASE LABEL =

-11.39 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 10781, CASE LABEL =

.22 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 7791, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 19.92 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	10.105115	10.088579	.0016	.2209
12 LNCONT	.947228	.953406	-.0065	-.7038
11 LNRMDIS	.350563	.352905	-.0067	-.1894

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM

COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.596994E-01	.596994E-01	-.111581E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

Vide KNHP 40 CH

```
  LESS THAN -4     3 :***
            -3.8   1 *
            -3.6   0
            -3.4   9 *********
            -3.2   1 *
            -3.0   1 *
            -2.8   2 **
            -2.6   3 ***
            -2.4   3 ***
            -2.2   2 **
            -2.0   1 *
            -1.8   3 ***
            -1.6  10 **********
            -1.4  19 *******************
            -1.2  14 **************
            -1.0  24 ************************
             -.8  56 ********************************************************
             -.6  97 ********************************************************************************
             -.4 117 ********************************************************************************
             -.2 124 ********************************************************************************
              .0 122 ********************************************************************************
              .2 103 ********************************************************************************
              .4  71 ***********************************************************************
              .6  31 ******************************
              .8  27 **************************
             1.0  37 ************************************
             1.2  38 *************************************
             1.4  23 ***********************
             1.6  11 ***********
             1.8   6 ******
             2.0   3 ***
             2.2   2 **
             2.4   0
             2.6   1 *
             2.8   0
             3.0   1 *
             3.2   0
             3.4   0
             3.6   0
             3.8   0
GREATER THAN 4     2 **
```

19.30 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 11227, CASE LABEL =

-11.39 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 10781, CASE LABEL =

.22 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 7791, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 19.92 PERCENT CONFIDENCE ELLIPSOID.

STATISTICS FOR 'BEST' SUBSET

SQUARED MULTIPLE CORRELATION	.93349
MULTIPLE CORRELATION	.96617
ADJUSTED SQUARED MULT. CORR.	.93335
RESIDUAL MEAN SQUARE	.059699
STANDARD ERROR OF EST.	.244335
F-STATISTIC	6772.04
NUMERATOR DEGREES OF FREEDOM	2
DENOMINATOR DEGREES OF FREEDOM	965
SIGNIFICANCE (TAIL PROB.)	.0000

NO	VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ
	INTERCEPT	10.1051	.0748617		10.677 134.98	.000		
12	LNCONT	.947228	.00877830		.904 107.91	.000	.981142	.80251
11	LNKMDIS	.350563	.0123599		.238 28.36	.000	.981142	.05544

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11
LNCONT	12	1.000	
LNKMDIS	11	-.137	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.05969937
AVERAGE DELETED RESIDUAL	.0001
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.05993579
SERIAL CORRELATION	.2752
DURBIN-WATSON STATISTIC	1.4485

HISTOGR OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BA OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.

NUMBER OF CASES READ. 11248
CASES WITH USE SET TO ZERO 10280
REMAINING NUMBER OF CASES 968

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	.75159	.90364	1.202309	.00000	3.95124	-.83
11 LNKMDIS	6.07345	.64179	.105671	3.52636	7.05012	-3.97
13 LNDEP	12.94617	.94644	.073106	10.95431	16.16803	-2.10

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

CORRELATIONS

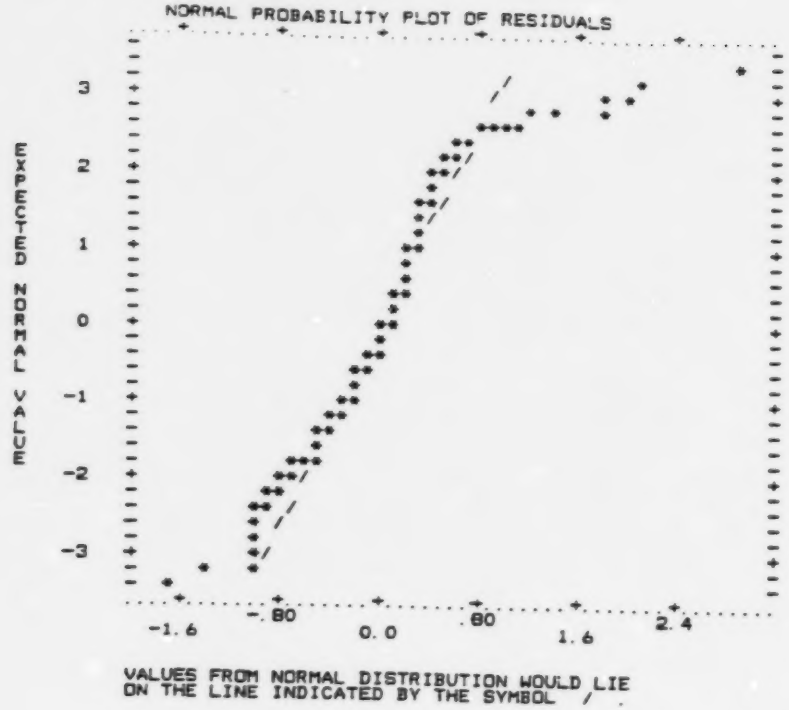
		LNCONT 12	LNKMDIS 11	LNDEP 13
LNCONT	12	1.000		
LNKMDIS	11	.137	1.000	
LNDEP	13	.937	.362	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

12 LNCONT *
13 LNDEP **
11 LNKMDIS .+*

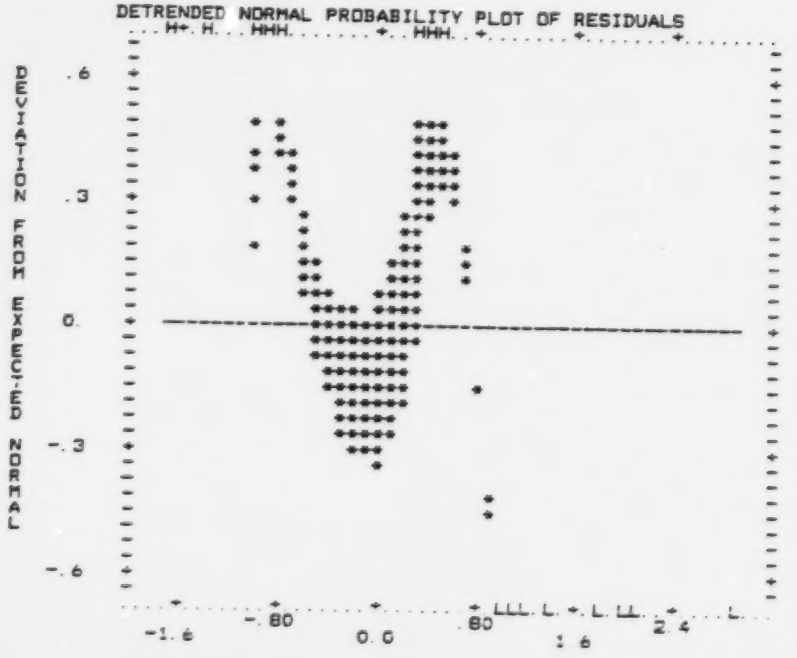
THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

	LESS THAN OR EQUAL TO		.117
.	.117	TO AND INCLUDING	.234
-	.234	TO AND INCLUDING	.351
+	.351	TO AND INCLUDING	.469
X	.469	TO AND INCLUDING	.586
N	.586	TO AND INCLUDING	.703
O	.703	TO AND INCLUDING	.820
*		GREATER THAN	.820

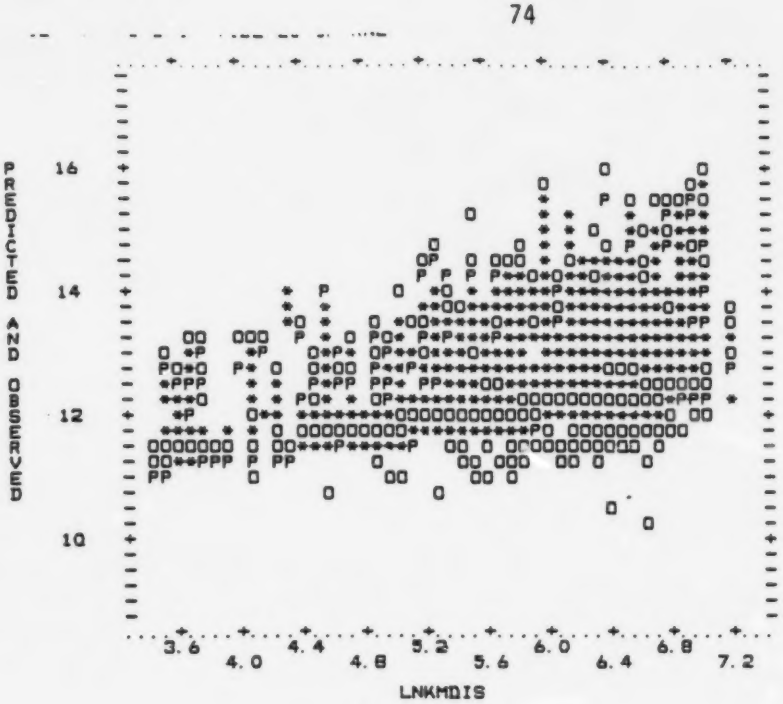


VALUES FROM NORMAL DISTRIBUTION WOULD LIE ON THE LINE INDICATED BY THE SYMBOL /.

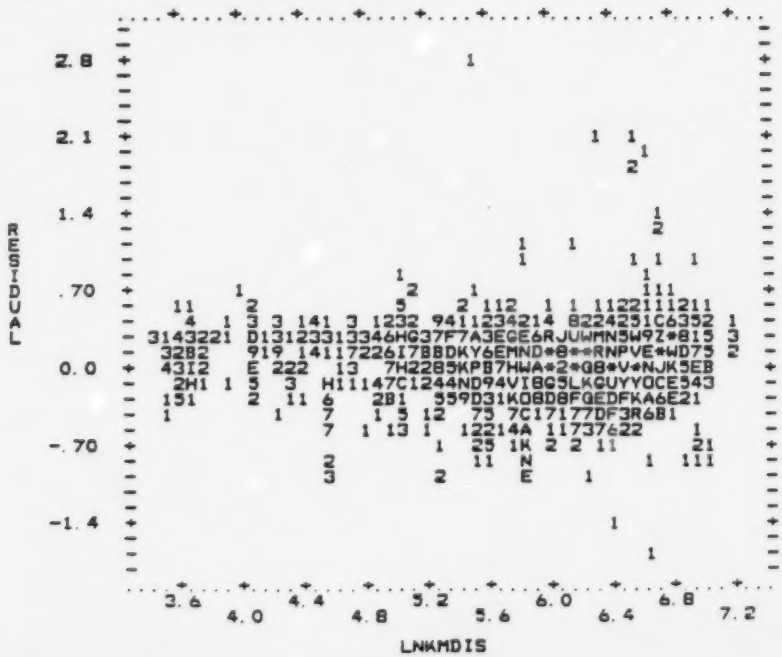
PAGE 18 BMDP1R VIDE 20 CH RNEP



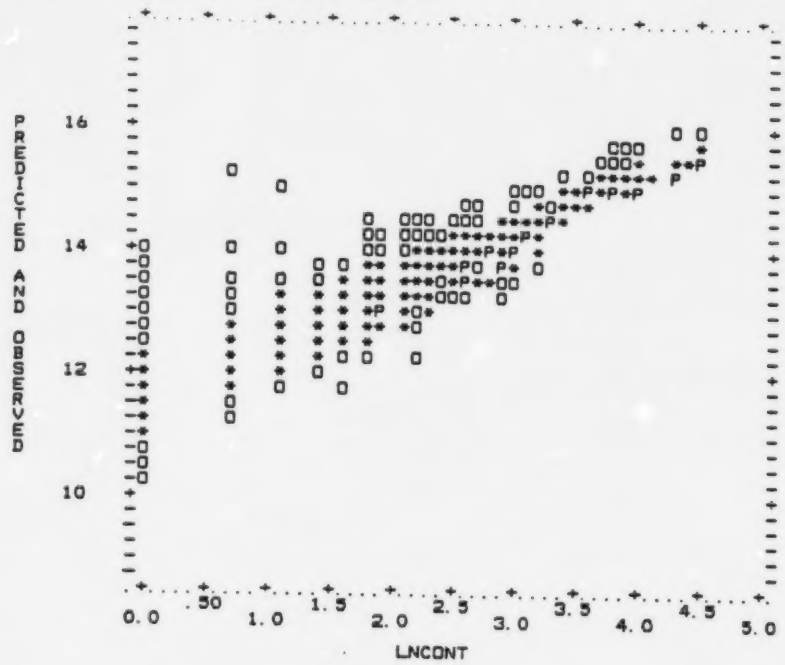
VALUES FROM NORMAL DISTRIBUTION WOULD LIE



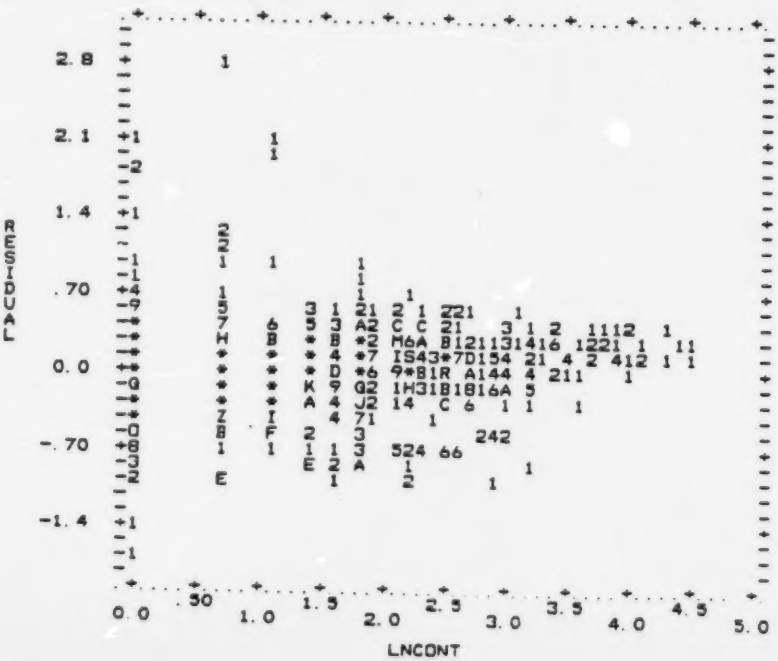
PAGE 12 BMDP1R VIDE 20 CH RNEP

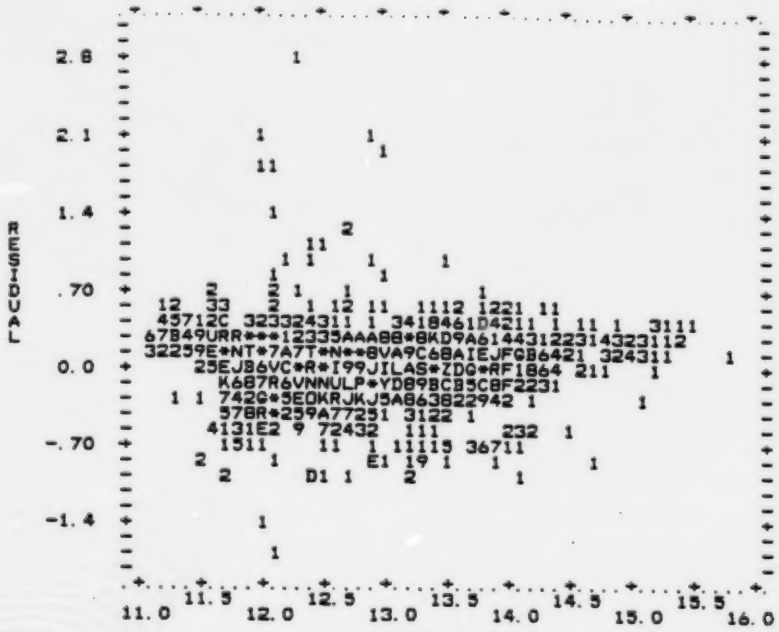


PAGE 9 BMDP1R VIDE 20 CH RNEP

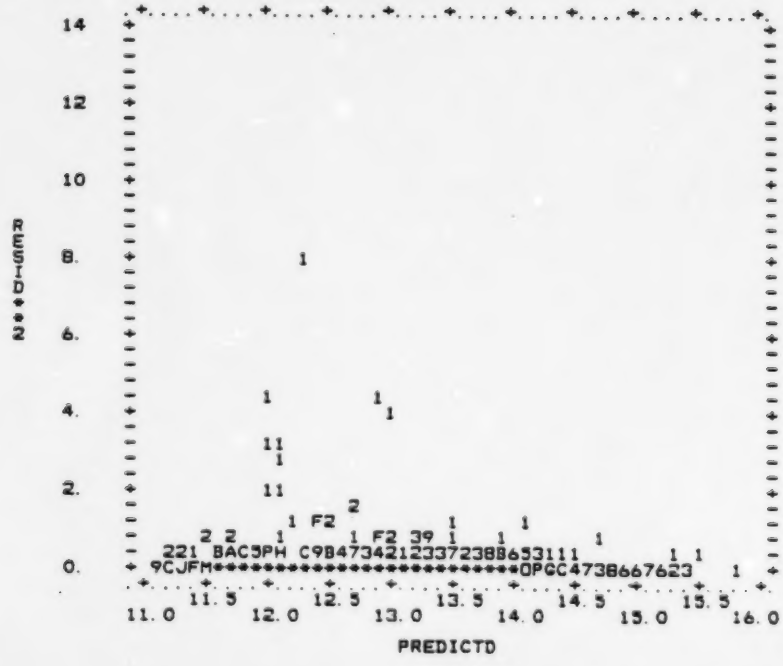


PAGE 10 BMDP1R VIDE 20 CH RNEP





PAGE 8 BMDP1R VIDE 20 CH RNEP



```
            3.0     0
            3.2     2 **
            3.4     2 **
            3.6     1 *
            3.8     1 *
GREATER THAN 4      9 *********
```

14.87 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 532, CASE LABEL =

9.41 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 4536, CASE LABEL =

.02 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 647, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 1.50 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	10.036341	10.038014	-.0002	-.0381
12 LNCONT	.799057	.800049	-.0012	-.1594
11 LNKMDIS	.308382	.307796	.0019	.0793

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM		
COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.918717E-01	.918717E-01	-.338366E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	.0000
AVERAGE RESIDUAL	.09187174
RESIDUAL MEAN SQUARE	.0000
AVERAGE DELETED RESIDUAL	
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.09195739
SERIAL CORRELATION	.3357
DURBIN-WATSON STATISTIC	1.3283

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

LESS THAN -4	2	**
-3.6	13	*************
-3.4	4	****
-3.2	3	***
-3.0	16	****************
-2.8	12	************
-2.6	14	**************
-2.4	19	*******************
-2.2	12	************
-2.0	25	*************************
-1.8	37	************************************
-1.6	57	**
-1.4	76	**
-1.2	82	**
-1.0	136	**
-.8	125	**
-.6	219	**
-.4	181	**
-.2	282	**
.0	232	**
.2	306	**
.4	222	**
.6	249	**
.8	238	**
1.0	150	**
1.2	70	**********************
1.4	23	**********
1.6	20	*********
1.8	10	**********
2.0	2	**
2.2	3	***
2.4	2	**
2.6	0	
2.8	2	**
3.0	0	
3.2	2	**
3.4	2	**
3.6	1	*
3.8	1	*
GREATER THAN 4	9	*********

14.87 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 532, CASE LABEL =

PAGE 6 BMDP9R VIDE RNEP 20 PIEDS CH

STATISTICS FOR 'BEST' SUBSET

SQUARED MULTIPLE CORRELATION	.87367
MULTIPLE CORRELATION	.93470
ADJUSTED SQUARED MULT. CORR.	.87358
RESIDUAL MEAN SQUARE	.091872
STANDARD ERROR OF EST.	.303104
F-STATISTIC	9875.93
NUMERATOR DEGREES OF FREEDOM	2
DENOMINATOR DEGREES OF FREEDOM	2856
SIGNIFICANCE (TAIL PROB.)	.0000

VARIABLE NO.	NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ
	INTERCEPT	10.0363	.0439196	11.773	228.52	.000		
12	LNCONT	.799057	.00622027	.860	128.46	.000	.986523	.72992
11	LNKMDIS	.308382	.00739373	.279	41.71	.000	.986523	.07695

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

PAGE 7 BMDP9R VIDE RNEP 20 PIEDS

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11
LNCONT	12	1.000	
LNKMDIS	11	-.116	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.09187174
AVERAGE DELETED RESIDUAL	.0000
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.09195739
SERIAL CORRELATION	.3357
DURBIN-WATSON STATISTIC	1.3283

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH ...N OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT

NUMBER OF CASES READ 11248
CASES WITH USE SET TO ZERO 8389
REMAINING NUMBER OF CASES 2859

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	1.08925	.91769	.842502	.00000	4.49981	-1.19
11 LNKMDIS	5.92606	.77205	.130280	3.43399	7.22548	-3.23
13 LNDEP	12.73420	.85249	.066945	10.35137	15.99304	-2.80

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

CORRELATIONS

		LNCONT 12	LNKMDIS 11	LNDEP 13
LNCONT	12	1.000		
LNKMDIS	11	.116	1.000	
LNDEP	13	.893	.379	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

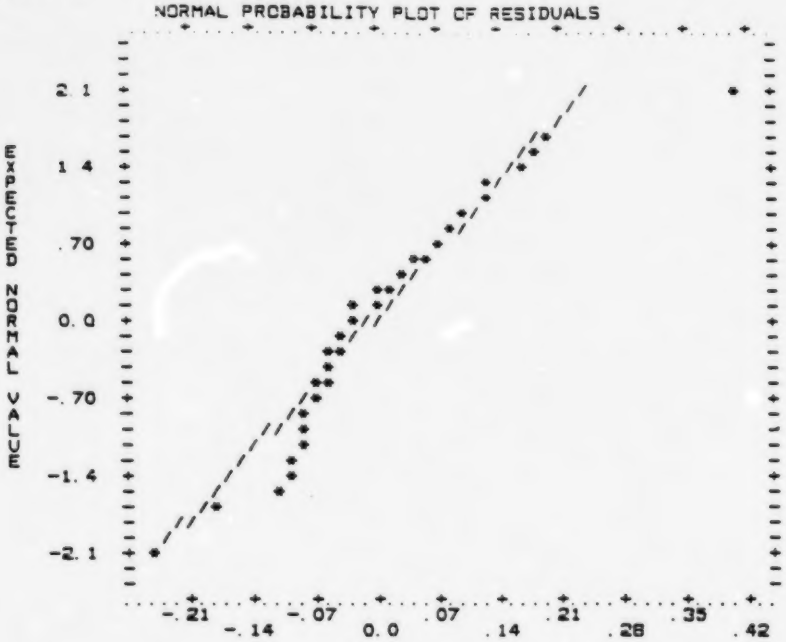
```
12 LNCONT    *
13 LNDEP     **
11 LNKMDIS   .**
```

THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

	LESS THAN OR EQUAL TO	.112
.	.112 TO AND INCLUDING	.223
-	.223 TO AND INCLUDING	.335
+	.335 TO AND INCLUDING	.446
X	.446 TO AND INCLUDING	.558
N	.558 TO AND INCLUDING	.669
O	.669 TO AND INCLUDING	.781
*	GREATER THAN	.781

PAGE 20 BMDP1R LCL 40 CH RNEP

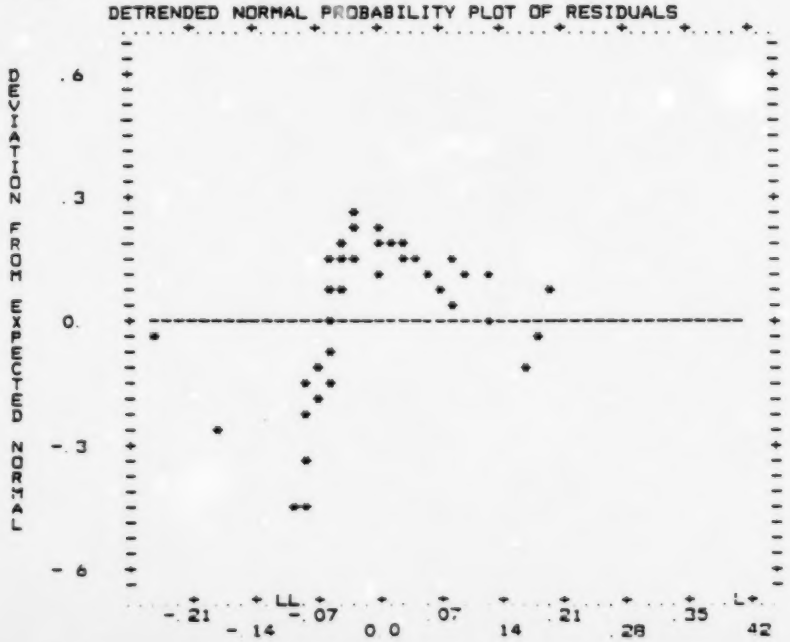
NORMAL PROBABILITY PLOT OF RESIDUALS

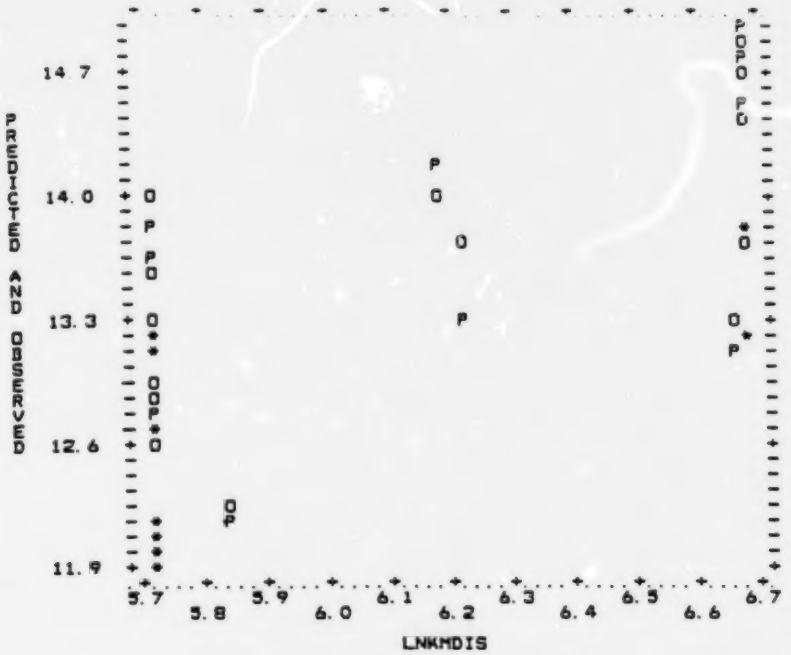


VALUES FROM NORMAL DISTRIBUTION WOULD LIE ON THE LINE INDICATED BY THE SYMBOL / .

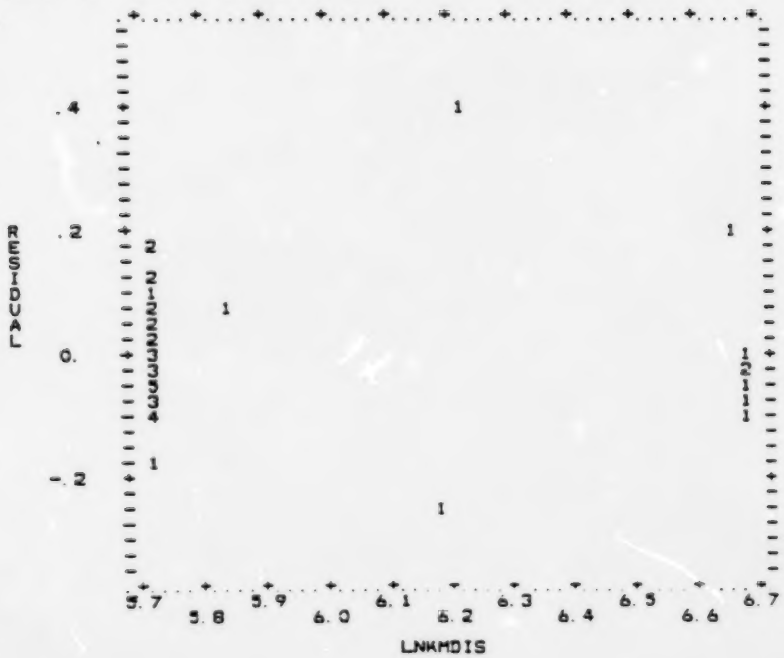
PAGE 21 BMDP1R LCL 40 CH RNEP

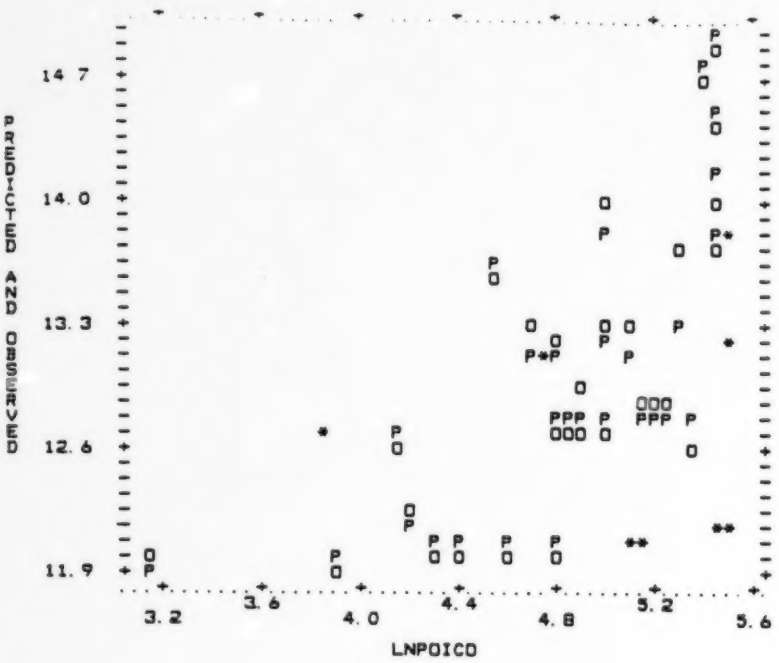
DETRENDED NORMAL PROBABILITY PLOT OF RESIDUALS



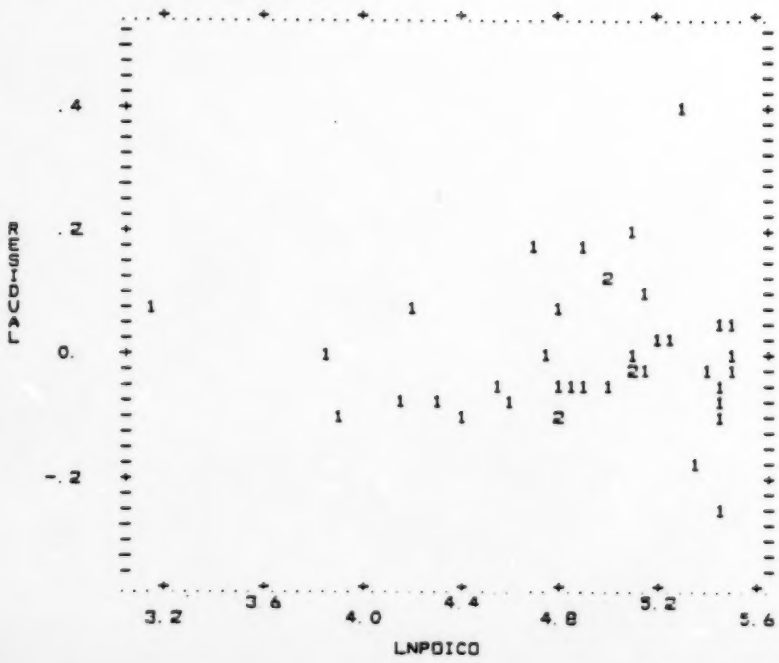


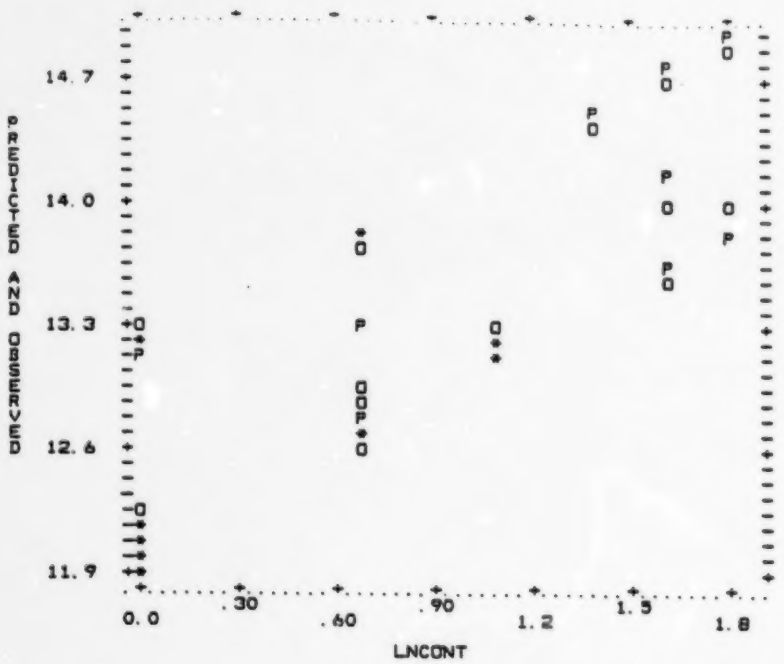
PAGE 14 BMDP1R LCL 40 CH RNEP



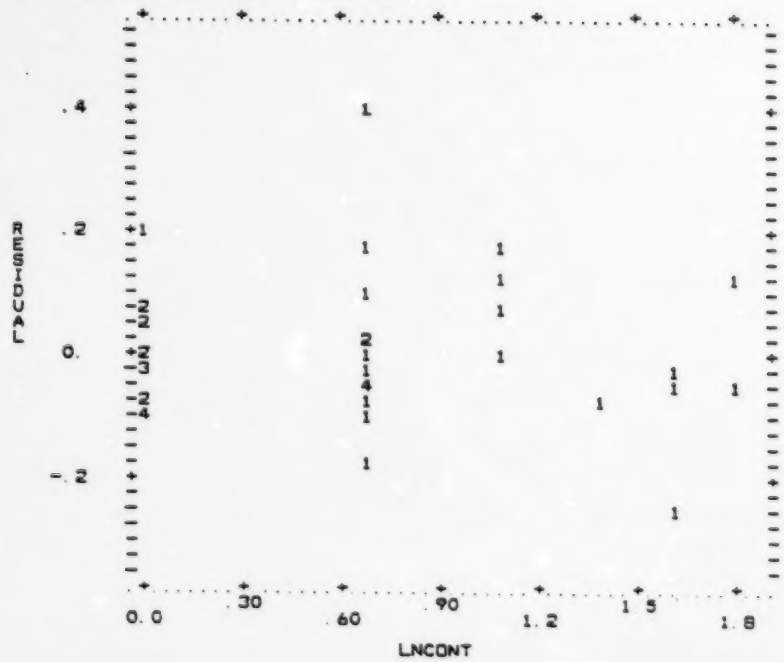


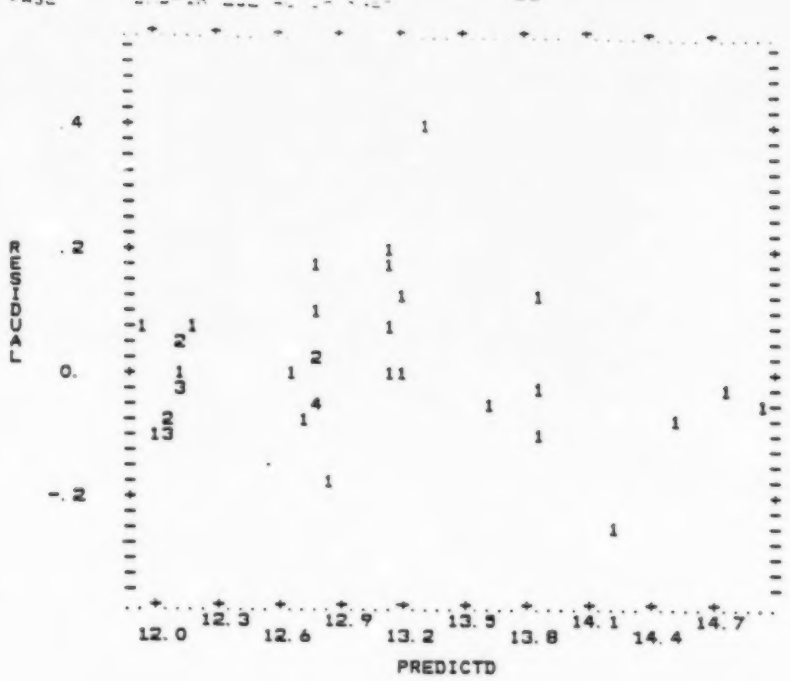
PAGE 12 BMDP1R LCL 40 CH RNEP



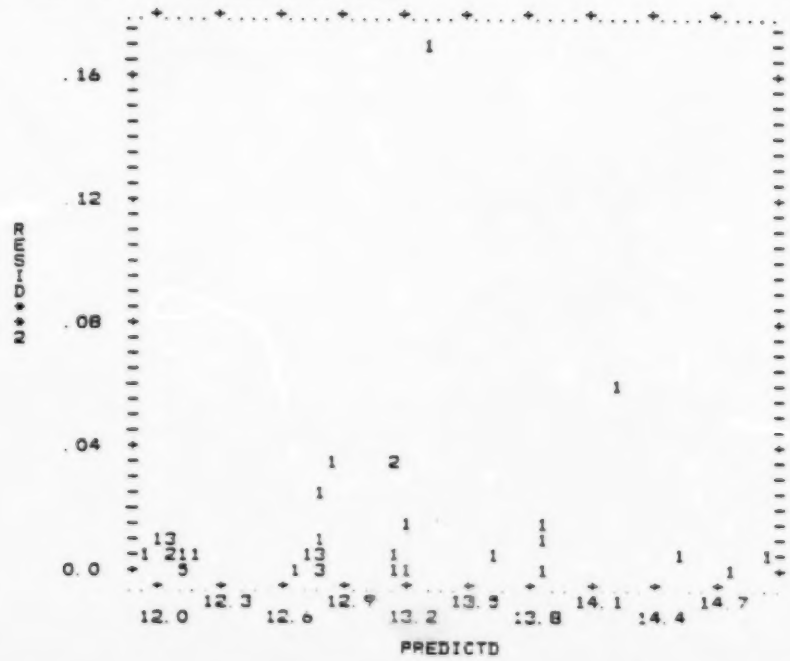


PAGE 10 BMDP1R LCL 40 CH RNEP





PAGE 8 BMDP1R LCL 40 CH RNEP



POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER

3.48 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 2060, CASE LABEL =

.20 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 1644, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 10.77 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	5.521061	5.744352	-.0404	-.6989
12 LNCONT	.982593	.999910	-.0176	-.4980
11 LNKMDIS	1.069623	1.024093	.0426	.7528
14 LNPOICO	.091475	.097539	-.0663	-.1436

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM		
COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.145722E-01	.145722E-01	.201897E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

LCL / ANEP / 40'

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11	LNPOICO 14
LNCONT	12	1.000		
LNKMDIS	11	-.175	1.000	
LNPOICO	14	-.159	-.444	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.01457223
AVERAGE DELETED RESIDUAL	.0006
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.01591046
SERIAL CORRELATION	-.0501
DURBIN-WATSON STATISTIC	1.9596

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

```
-2.2    1 *
-2.0    0
-1.8    1 *
-1.6    0
-1.4    0
-1.2    0
-1.0    3 ***
 -.8    6 ******
 -.6    5 *****
 -.4    5 *****
 -.2    2 **
  .0    3 ***
  .2    2 **
  .4    2 **
  .6    2 **
  .8    2 **
 1.0    2 **
 1.2    1 *
 1.4    1 *
 1.6    1 *
 1.8    0
 2.0    0
 2.2    0
 2.4    0
 2.6    0
 2.8    0
 3.0    0
 3.2    0
 3.4    1 *
```

STATISTICS FOR 'BEST' SUBSET

MALLOWS' CP	4.00
SQUARED MULTIPLE CORRELATION	.97977
MULTIPLE CORRELATION	.98984
ADJUSTED SQUARED MULT. CORR.	.97809
RESIDUAL MEAN SQUARE	.014572
STANDARD ERROR OF EST.	.120716
F-STATISTIC	581.28
NUMERATOR DEGREES OF FREEDOM	3
DENOMINATOR DEGREES OF FREEDOM	36
SIGNIFICANCE (TAIL PROB.)	.0000

NOTE THAT THE ABOVE F-STATISTIC AND ASSOCIATED SIGNIFICANCE TEND TO BE LIBERAL WHENEVER A SUBSET OF VARIABLES IS SELECTED BY THE CP OR ADJUSTED R-SQUARED CRITERIA.

NO.	VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ
	INTERCEPT	5.52106	.319498	6.770	17.28	.000		
12	LNCONT	.982593	.0347696	.706	28.26	.000	.900119	.44871
11	LNKMDIS	1.06962	.0604817	.487	17.69	.000	.741642	.17573
14	LNPOICO	.0914751	.0422352	.059	2.17	.037	.745680	.00264

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

COVARIANCES OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11	LNPOICO 14
LNCONT	12	.001209		
LNKMDIS	11	-.000367	.003658	
LNPOICO	14	-.000233	-.001133	.001784

PAGE 4 BMDP9R LCL RNEP 40 PIEDS

NUMBER OF CASES READ 11249
CASES WITH USE SET TO ZERO 11209
REMAINING NUMBER OF CASES 40

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	.59742	.58598	.980853	.00000	1.79176	-1.02
11 LNKMDIS	5.91796	.37112	.062710	5.72359	6.68336	.52
14 LNPOICO	4.90667	.53000	.108017	3.13549	5.49717	-3.34
13 LNDEP	12.88690	.81549	.063281	11.90091	14.86619	-1.21

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

PAGE 5 BMDP9R LCL RNEP 40 PIEDS

CORRELATIONS

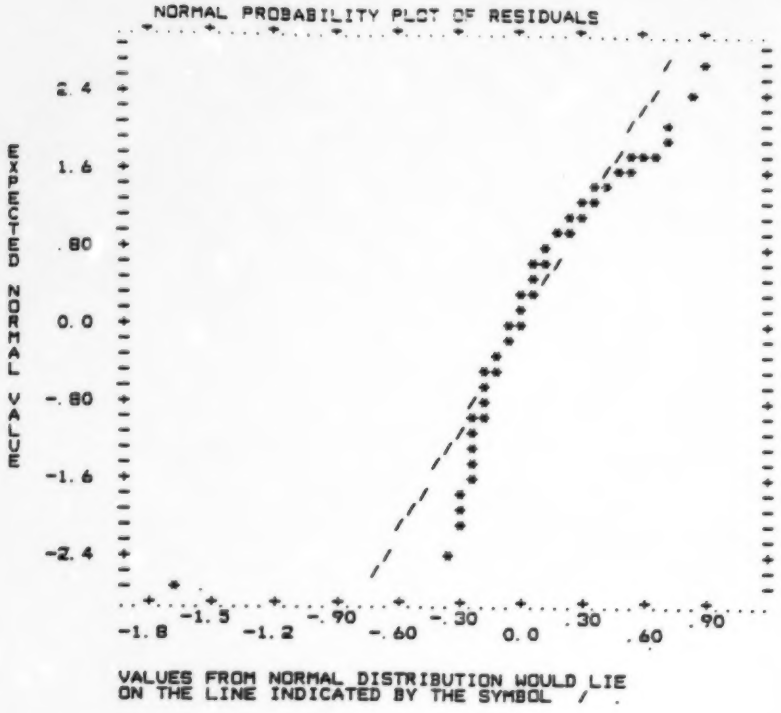
		LNCONT 12	LNKMDIS 11	LNPOICO 14	LNDEP 13
LNCONT	12	1.000			
LNKMDIS	11	.277	1.000		
LNPOICO	14	.268	.485	1.000	
LNDEP	13	.857	.711	.484	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

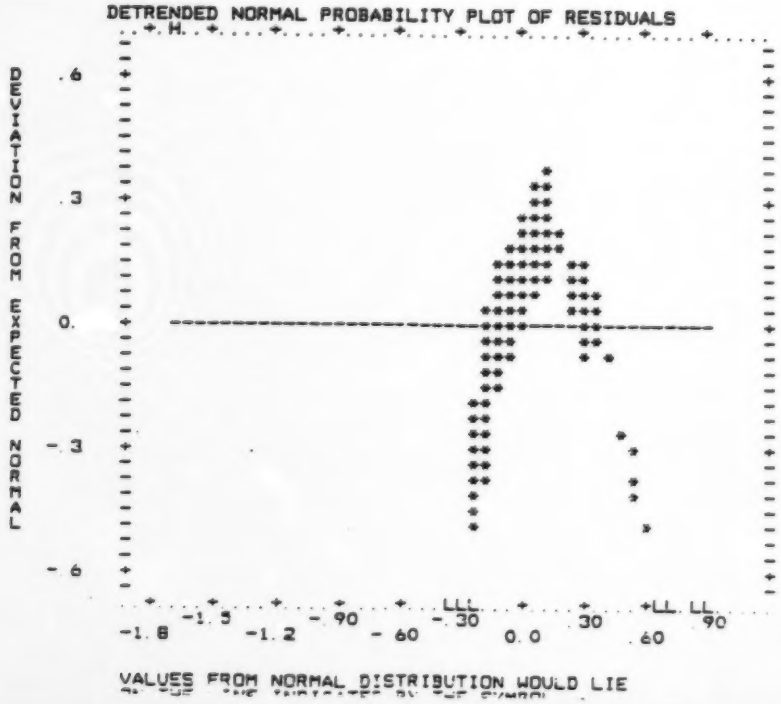
12 LNCONT *
13 LNDEP **
11 LNKMDIS -0*
14 LNPOICO -XX*

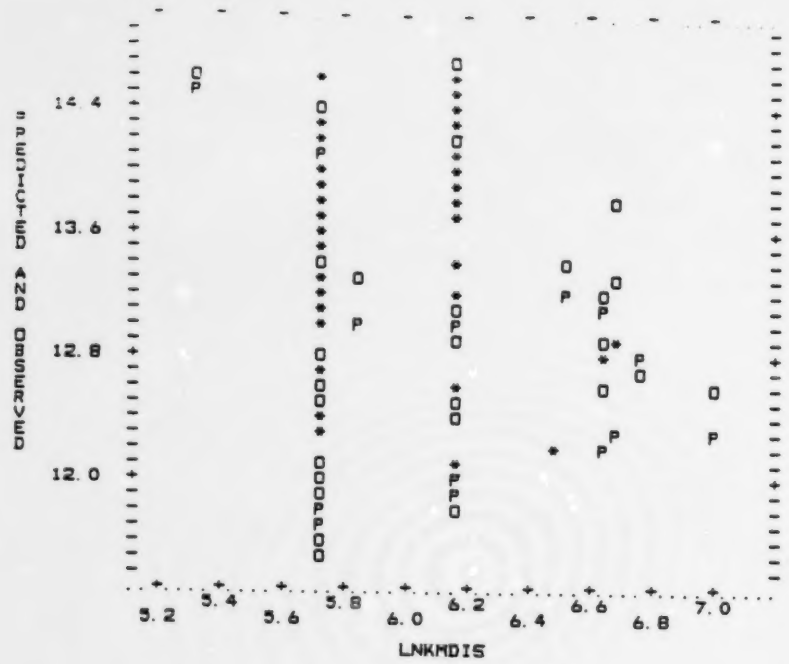
THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

	LESS THAN OR EQUAL TO		.107
.	.107	TO AND INCLUDING	.214
-	.214	TO AND INCLUDING	.321
+	.321	TO AND INCLUDING	.428
X	.428	TO AND INCLUDING	.535
N	.535	TO AND INCLUDING	.643
O	.643	TO AND INCLUDING	.750
*		GREATER THAN	.750

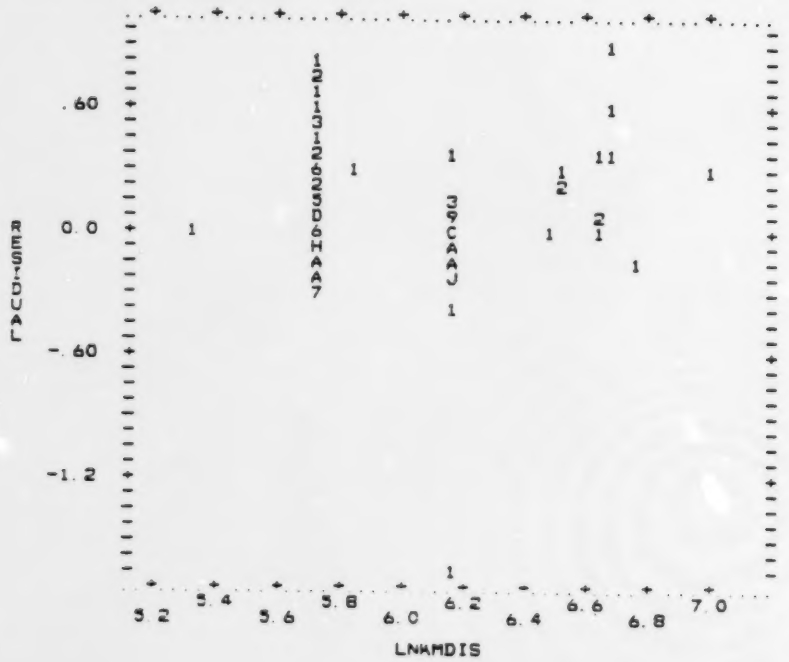


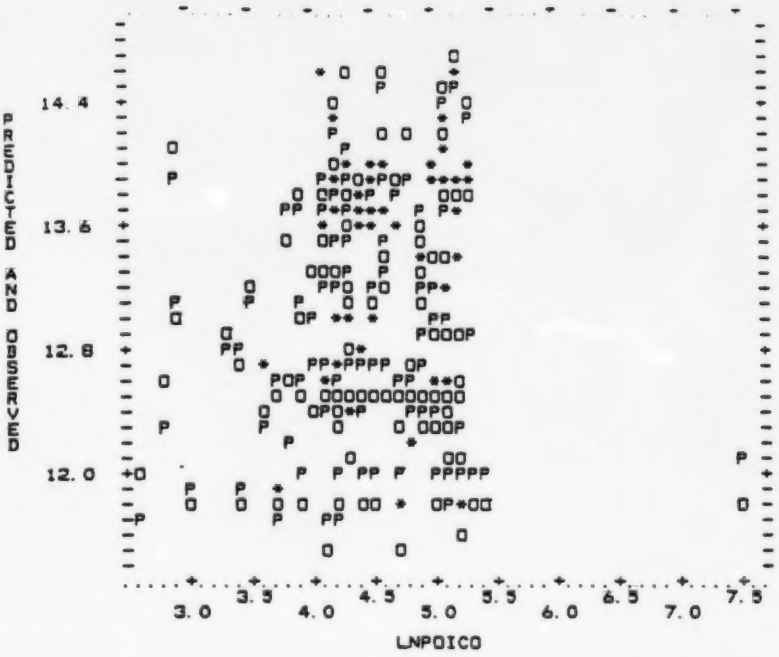
PAGE 21 BMDP1R LCL 20 CH RNEP



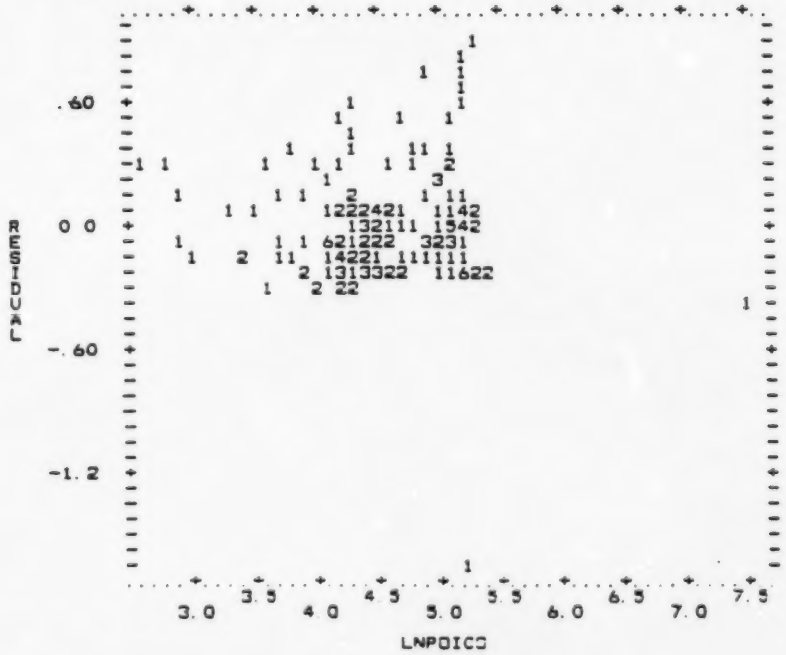


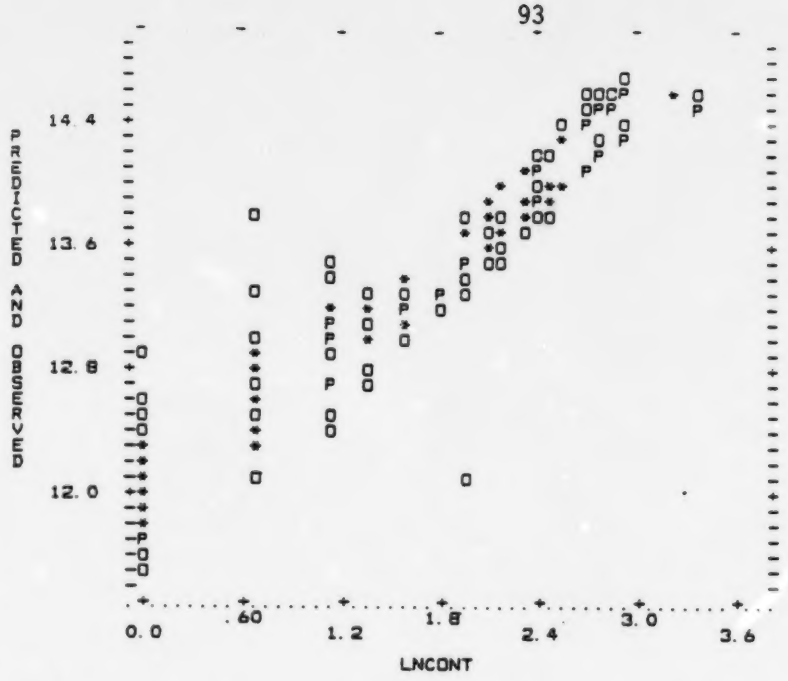
PAGE 14 BMDP1R LCL 20 CH RNEP



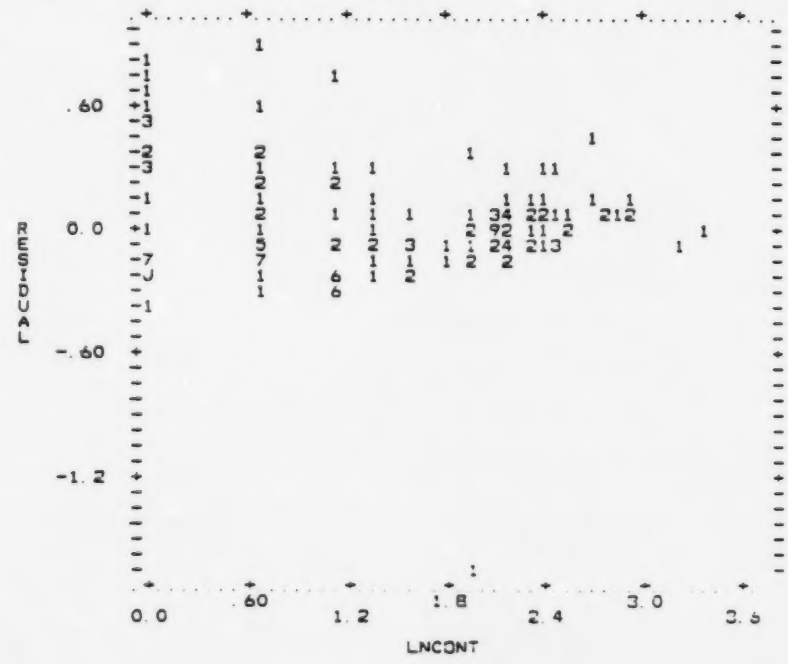


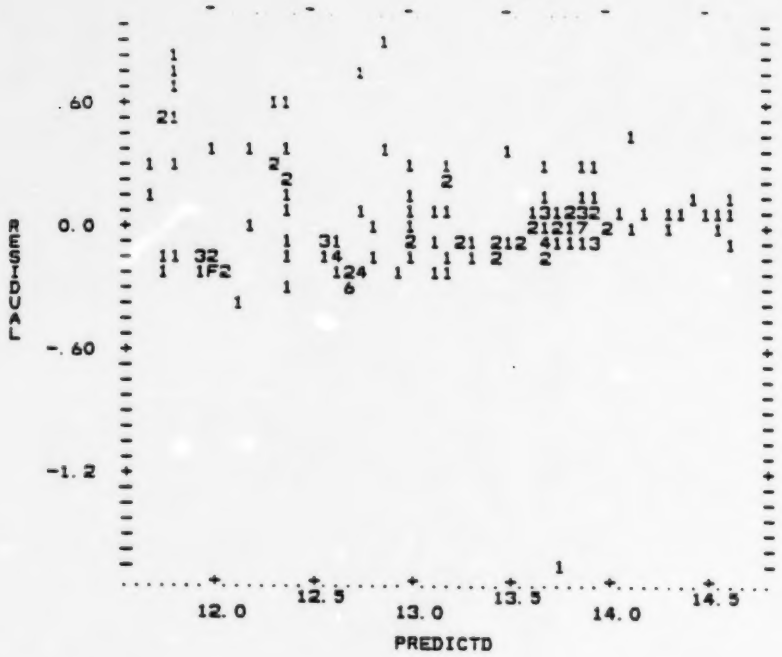
PAGE 12 BMDP1R LCL 20 CH RNEP



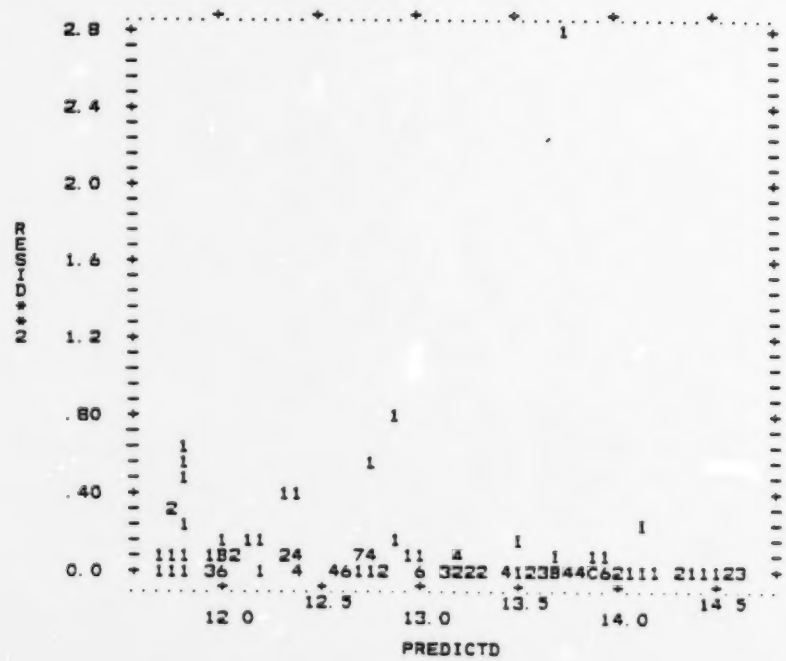


PAGE 10 BMDP1R LCL 20 CH RNEP





PAGE 8 BMDP1R LCL 20 CH RNEP



CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 1495, CASE LABEL

.18 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES
WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 1495, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD
MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 9.18
PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	8.550301	8.310175	.0281	.5096
12 LNCONT	.884411	.894368	-.0113	-.4271
11 LNKMDIS	.519576	.547649	-.0540	-.3606
14 LNPOICO	.051563	.066838	-.2962	-.4357

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND
RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM		
COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.757197E-01	.757197E-01	-.316705E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR
A VALUE OUT OF RANGE.

LCL RNCP 20'

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11	LNPOICO 14
LNCONT	12	1.000		
LNKMDIS	11	.323	1.000	
LNPOICO	14	-.056	-.198	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.07571965
AVERAGE DELETED RESIDUAL	.0010
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.07839989
SERIAL CORRELATION	.1758
DURBIN-WATSON STATISTIC	1.6433

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

LESS THAN -4	1	:*
-1.4	2	**
-1.2	4	****
-1.0	18	******************
-.8	20	********************
-.6	15	***************
-.4	17	*****************
-.2	22	**********************
.0	15	***************
.2	16	****************
.4	8	********
.6	2	**
.8	5	*****
1.0	5	*****
1.2	5	*****
1.4	2	**
1.6	1	*
1.8	2	**
2.0	1	*
2.2	2	**
2.4	1	*
2.6	2	**
2.8	0	
3.0	1	*
3.2	1	*

18 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 1381, CASE LABEL =

STATISTICS FOR 'BEST' SUBSET

MALLOWS' CP	4.00
SQUARED MULTIPLE CORRELATION	.90025
MULTIPLE CORRELATION	.94881
ADJUSTED SQUARED MULT. CORR.	.89842
RESIDUAL MEAN SQUARE	.075720
STANDARD ERROR OF EST.	.275172
F-STATISTIC	493.36
NUMERATOR DEGREES OF FREEDOM	3
DENOMINATOR DEGREES OF FREEDOM	164
SIGNIFICANCE (TAIL PROB.)	.0000

NOTE THAT THE ABOVE F-STATISTIC AND ASSOCIATED SIGNIFICANCE TEND TO BE LIBERAL WHENEVER A SUBSET OF VARIABLES IS SELECTED BY THE CP OR ADJUSTED R-SQUARED CRITERIA.

NO.	VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOL-ERANCE	CONTRI-BUTION TO R-SQ
	INTERCEPT	8.55030	.471192	9.903	18.15	.000		
12	LNCONT	.884411	.0233146	.989	37.93	.000	.895582	.87525
11	LNKMDIS	.519576	.0778588	.177	6.67	.000	.863114	.02709
14	LNPOICO	.0515634	.0350532	.037	1.47	.143	.960671	.00132

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

COVARIANCES OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11	LNPOICO 14
LNCONT	12	.0005436		
LNKMDIS	11	.0005864	-.0060620	
LNPOICO	14	-.0000456	-.0005407	.0012287

PAGE 4 BMDP9R LCL RNEP 20 PIEDS

NUMBER OF CASES READ. 11249
CASES WITH USE SET TO ZERO 11081
REMAINING NUMBER OF CASES 168

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	1.29261	.96508	.746617	.00000	3.36730	-1.34
11 LNKMDIS	5.96912	.29438	.049317	5.32788	6.99393	-2.18
14 LNPOICO	4.57012	.61977	.135614	2.56495	7.54433	-3.24
13 LNDEP	13.03056	.86339	.066259	11.51192	14.74657	-1.76

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

PAGE 5 BMDP9R LCL RNEP 20 PIEDS

CORRELATIONS

		LNCONT 12	LNKMDIS 11	LNPOICO 14	LNDEP 13
LNCONT	12	1.000			
LNKMDIS	11	-.319	1.000		
LNPOICO	14	-.009	.191	1.000	
LNDEP	13	.932	-.131	.062	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

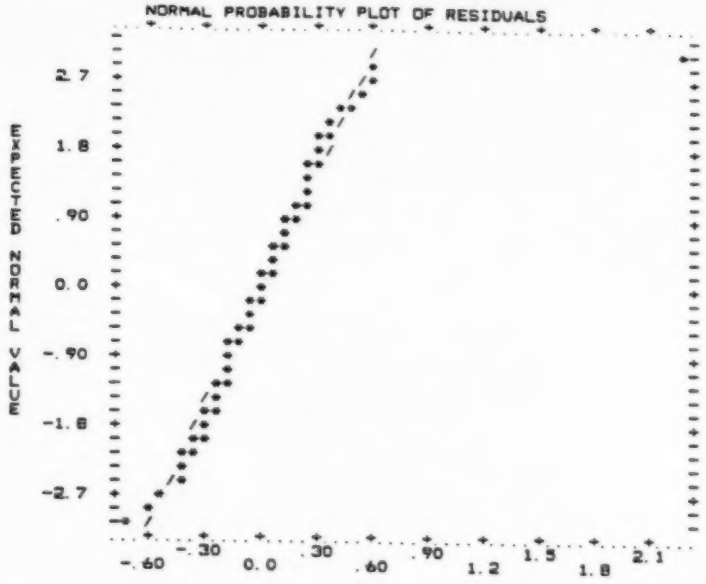
```
12 LNCONT    *
13 LNDEP     **
11 LNKMDIS   -.*
14 LNPOICO     .*
```

THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

	LESS THAN OR EQUAL TO		
.	.116	TO AND INCLUDING	.116
-	.233	TO AND INCLUDING	.349
+	.349	TO AND INCLUDING	.466
X	.466	TO AND INCLUDING	.582
N	.582	TO AND INCLUDING	.699
O	.699	TO AND INCLUDING	.815
*		GREATER THAN	.815

PAGE 20 BMDP1R FCL 40 CH RNEP

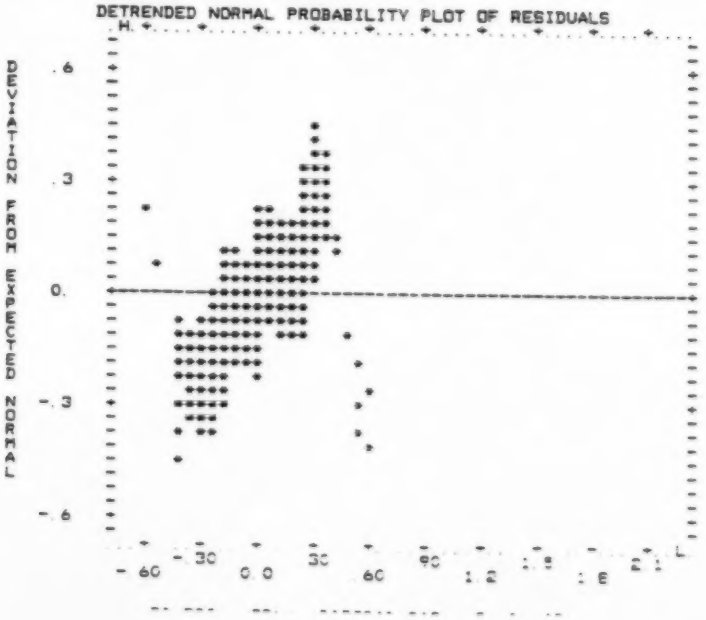
NORMAL PROBABILITY PLOT OF RESIDUALS

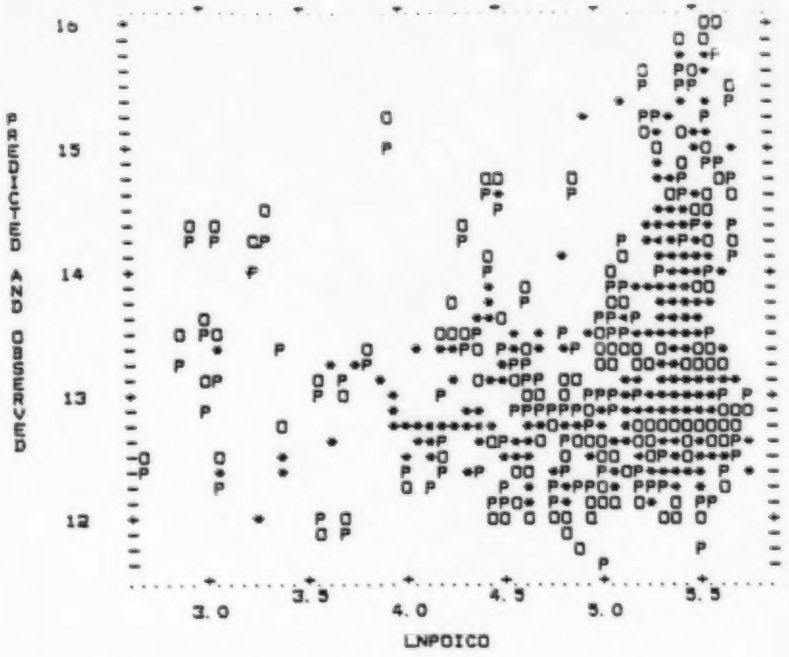


VALUES FROM NORMAL DISTRIBUTION WOULD LIE
ON THE LINE INDICATED BY THE SYMBOL /

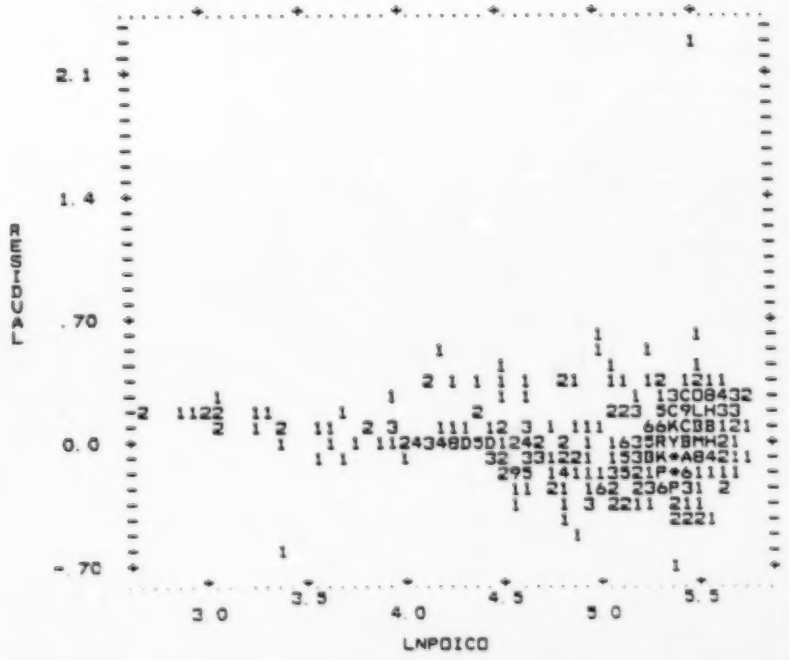
PAGE 21 BMDP1R FCL 40 CH RNEP

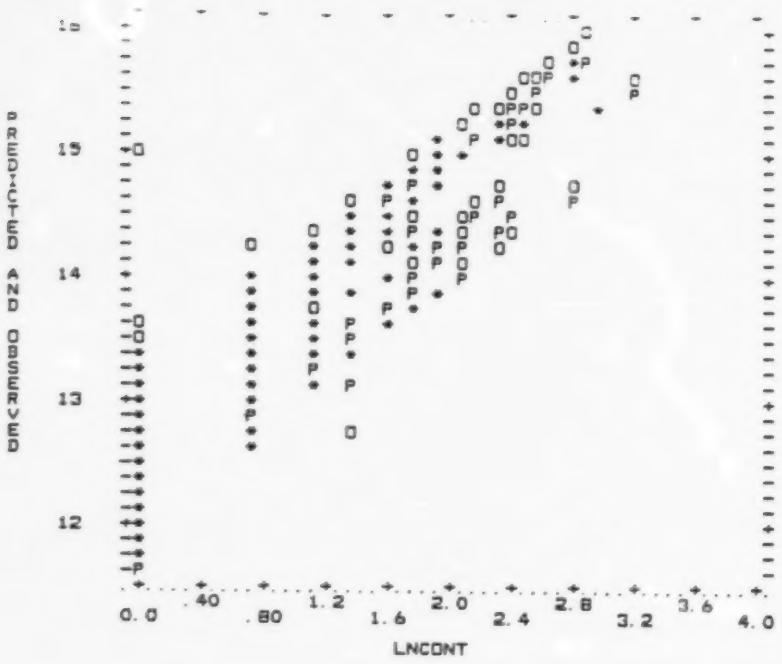
DETRENDED NORMAL PROBABILITY PLOT OF RESIDUALS



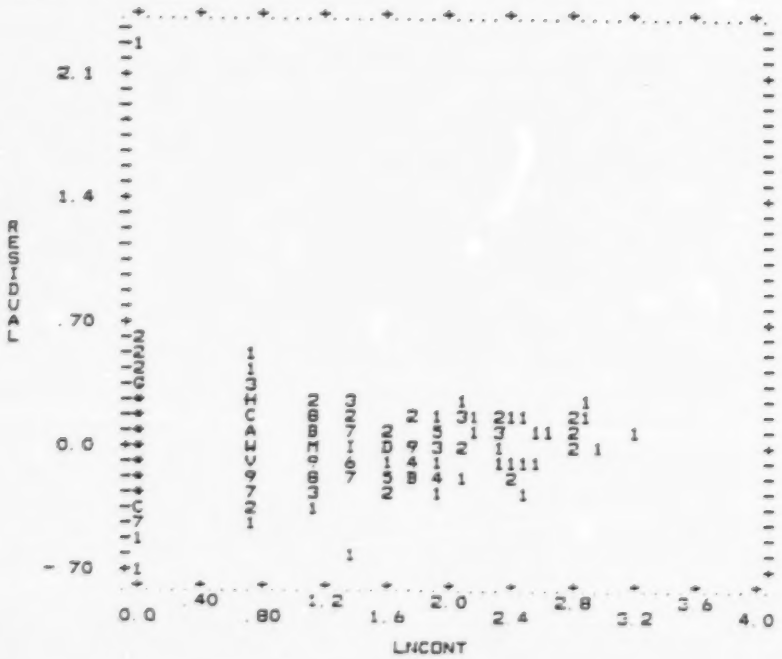


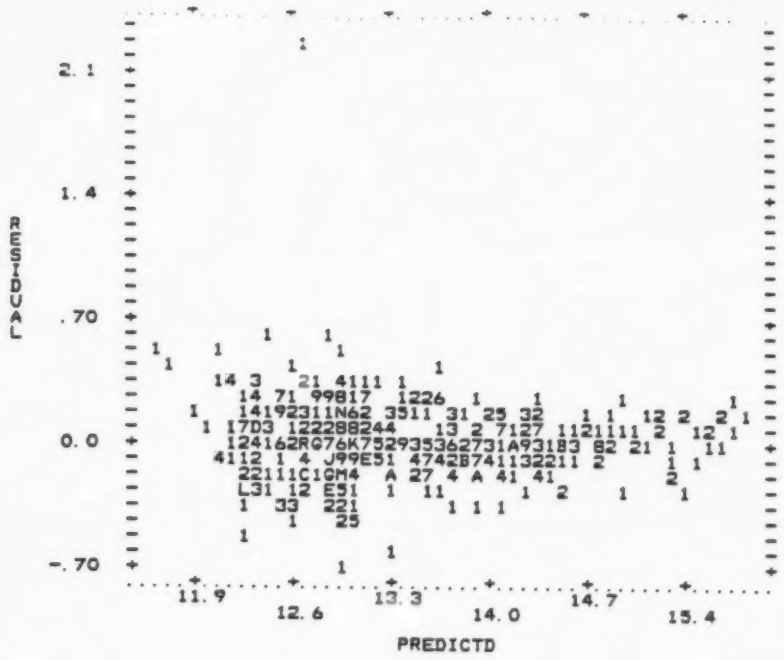
PAGE 12 BMDP1R FCL 40 CH RNEP



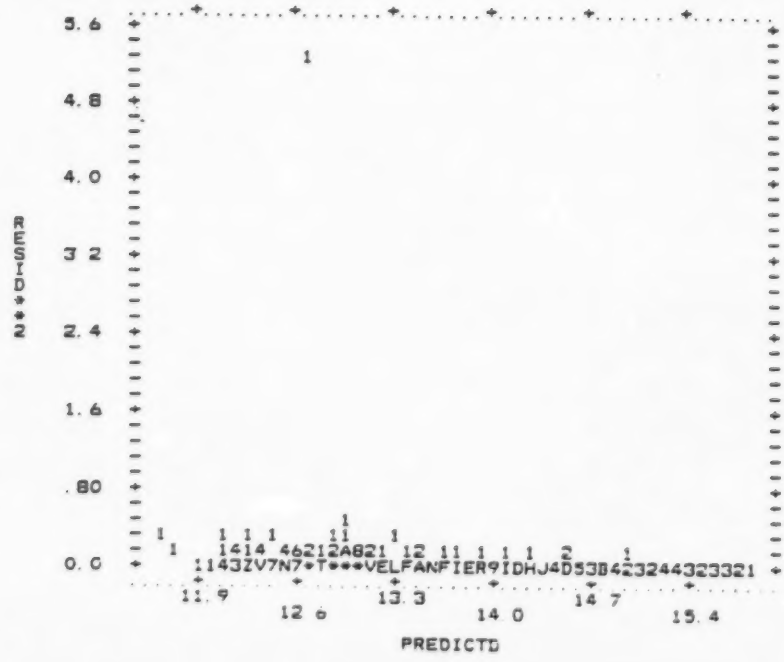


PAGE 10 BMDP1R FCL 40 CH RNEP





PAGE 8 BMDP1R FCL 40 CH RNEP



```
            3.0     2 **
            3.2     0
            3.4     0
            3.6     0
            3.8     0
GREATER THAN 4      1 *
```

22.25 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 870, CASE LABEL =

11.79 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 3728, CASE LABEL =

.11 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 3728, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 4.45 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	7.526895	7.520810	.0008	.0558
12 LNCONT	.942984	.945940	-.0031	-.3120
11 LNKMDIS	.673328	.677130	-.0056	-.2573
14 LNPOICO	.187232	.182850	.0234	.3332

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM		
COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.375648E-01	.375648E-01	.260083E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

SUMMARY STATISTICS FOR RESIDUALS FOR MODEL 40

(CASES WITH POSITIVE WEIGHT)

AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.03756479
AVERAGE DELETED RESIDUAL	.0001
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.03773090
SERIAL CORRELATION	.3948
DURBIN-WATSON STATISTIC	1.2083

HISTOGRAM OF STANDARDIZED (STUDENTIZED) RESIDUALS
EACH BIN OF THE HISTOGRAM IS LABELED WITH ITS LOWER LIMIT.
NOTE THAT IF THE COUNT FOR A BIN EXCEEDS 100, ONLY
100 ASTERISKS WILL BE PRINTED.

-3.8	1	*
-3.6	0	
-3.4	0	
-3.2	1	*
-3.0	1	*
-2.8	0	
-2.6	0	
-2.4	3	***
-2.2	5	*****
-2.0	9	*********
-1.8	3	***
-1.6	20	********************
-1.4	28	****************************
-1.2	34	**********************************
-1.0	43	***
-.8	60	**
-.6	53	***
-.4	58	**
-.2	84	**
.0	88	**
.2	50	**
.4	45	***
.6	45	***
.8	35	***********************************
1.0	43	***
1.2	35	***********************************
1.4	15	***************
1.6	11	***********
1.8	7	*******
2.0	1	*
2.2	2	**
2.4	0	
2.6	2	**
2.8	1	*
3.0	2	**
3.2	0	
3.4	0	
3.6	0	
3.8	0	
GREATER THAN 4	1	*

22.25 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 870, CASE LABEL =

STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG

STATISTICS FOR 'BEST' SUBSET

MALLOWS' CP	4.00
SQUARED MULTIPLE CORRELATION	.94198
MULTIPLE CORRELATION	.97056
ADJUSTED SQUARED MULT. CORR.	.94176
RESIDUAL MEAN SQUARE	.037565
STANDARD ERROR OF EST.	.193816
F-STATISTIC	4232.26
NUMERATOR DEGREES OF FREEDOM	3
DENOMINATOR DEGREES OF FREEDOM	782
SIGNIFICANCE (TAIL PROB.)	.0000

NOTE THAT THE ABOVE F-STATISTIC AND ASSOCIATED SIGNIFICANCE TEND TO BE LIBERAL WHENEVER A SUBSET OF VARIABLES IS SELECTED BY THE CP OR ADJUSTED R-SQUARED CRITERIA.

NO	VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ
	INTERCEPT	7.52690	.109088	9.372	69.00	.000		
12	LNCONT	.942984	.00947237	.858	99.55	.000	.998684	.73526
11	LNKMDIS	.673328	.0147753	.395	45.57	.000	.985392	.15407
14	LNPOICO	.187232	.0131525	.124	14.24	.000	.985058	.01503

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 12	LNKMDIS 11	LNPOICO 14
LNCONT	12	1.000		
LNKMDIS	11	-.020	1.000	
LNPOICO	14	-.027	-.119	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.03756479
AVERAGE DELETED RESIDUAL	.0001

PAGE 4 BMDP9R FCL RNEP 40 PIEDS

NUMBER OF CASES READ 11249
CASES WITH USE SET TO ZERO 10463
REMAINING NUMBER OF CASES 786

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
12 LNCONT	.53444	.73077	1.367356	.00000	3.21888	-.73
11 LNKMDIS	6.34710	.47164	.074309	4.77068	7.18007	-3.34
14 LNPOICO	5.14199	.52993	.103058	2.70805	5.73657	-4.59
13 LNDEP	13.26730	.80312	.060534	11.74879	15.98186	-1.87

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

PAGE 5 BMDP9R FCL RNEP 40 PIEDS

CORRELATIONS

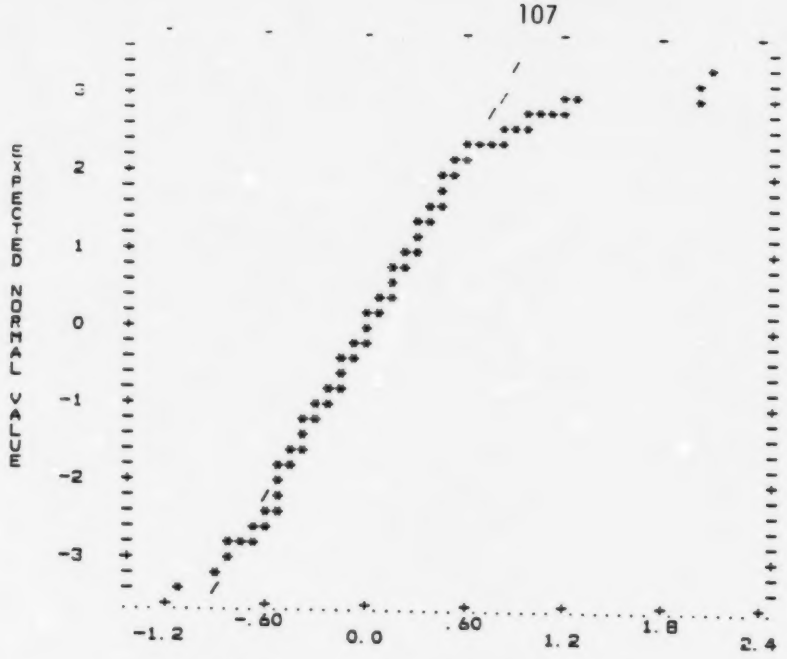
		LNCONT 12	LNKMDIS 11	LNPOICO 14	LNDEP 13
LNCONT	12	1.000			
LNKMDIS	11	.024	1.000		
LNPOICO	14	.030	.119	1.000	
LNDEP	13	.871	.431	.196	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

```
12 LNCONT    *
13 LNDEP     **
11 LNKMDIS    +*
14 LNPOICO    ..*
```

THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

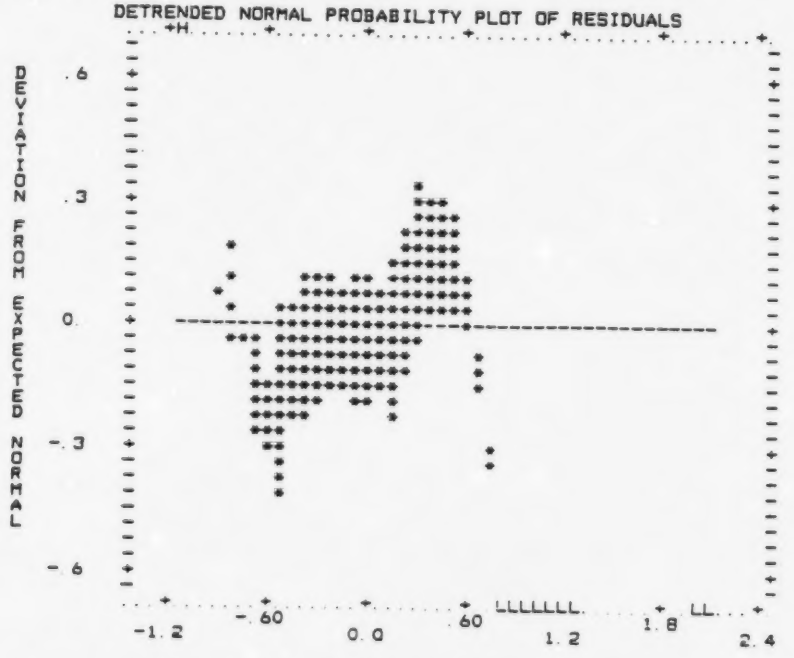
	LESS THAN OR EQUAL TO		.109
.	.109	TO AND INCLUDING	.218
-	.218	TO AND INCLUDING	.327
+	.327	TO AND INCLUDING	.436
X	.436	TO AND INCLUDING	.544
N	.544	TO AND INCLUDING	.653
O	.653	TO AND INCLUDING	.762
*		GREATER THAN	.762



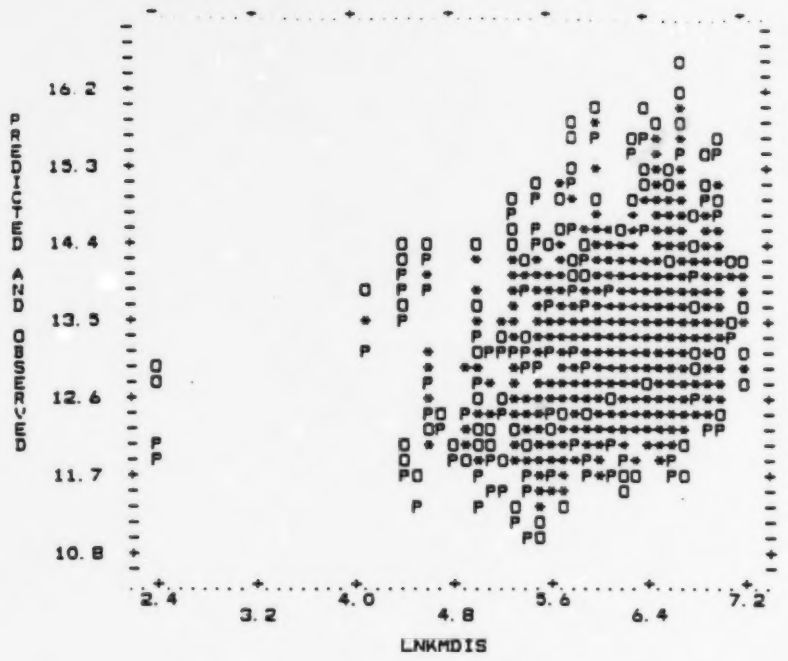
VALUES FROM NORMAL DISTRIBUTION WOULD LIE
ON THE LINE INDICATED BY THE SYMBOL /

PAGE 18 BMDP1R FCL 20 CH RNEP

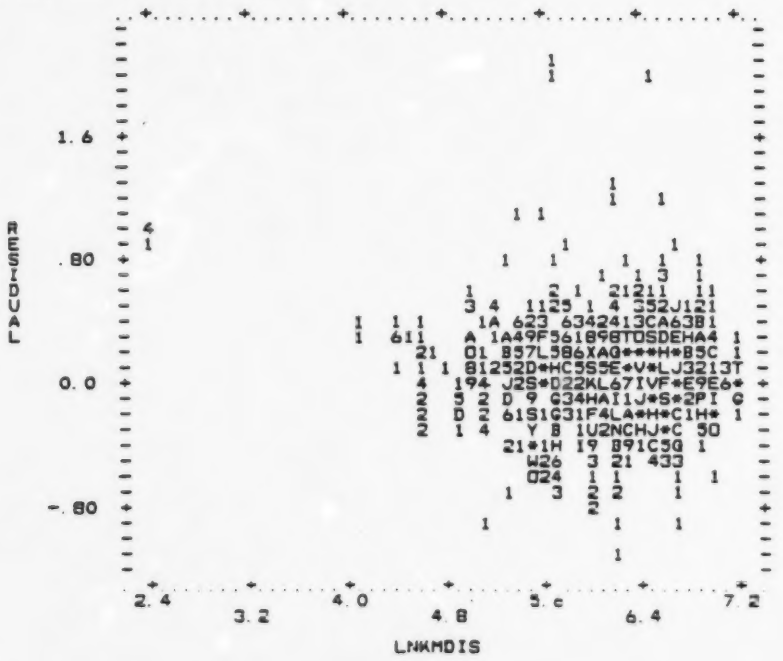
DETRENDED NORMAL PROBABILITY PLOT OF RESIDUALS

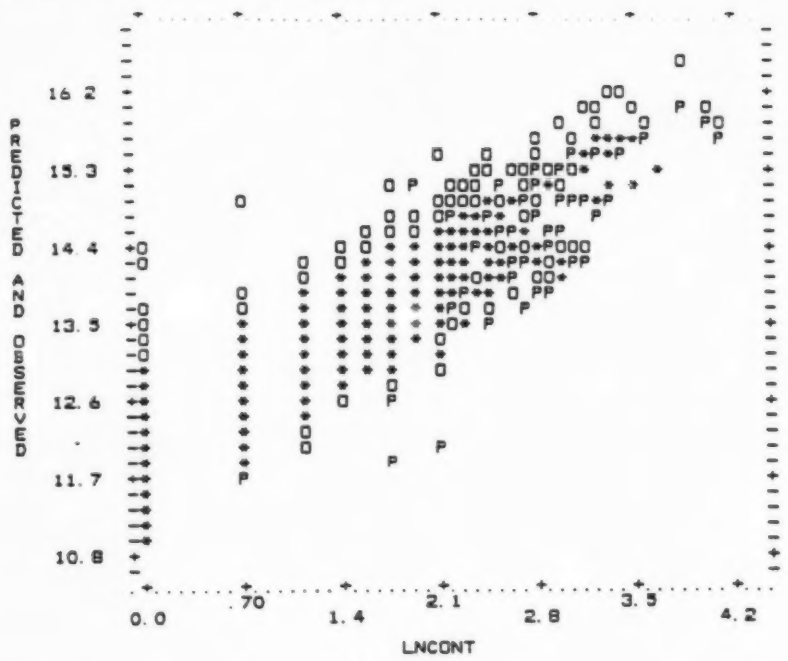


VALUES FROM NORMAL DISTRIBUTION WOULD LIE
ON THE LINE INDICATED BY THE SYMBOL -

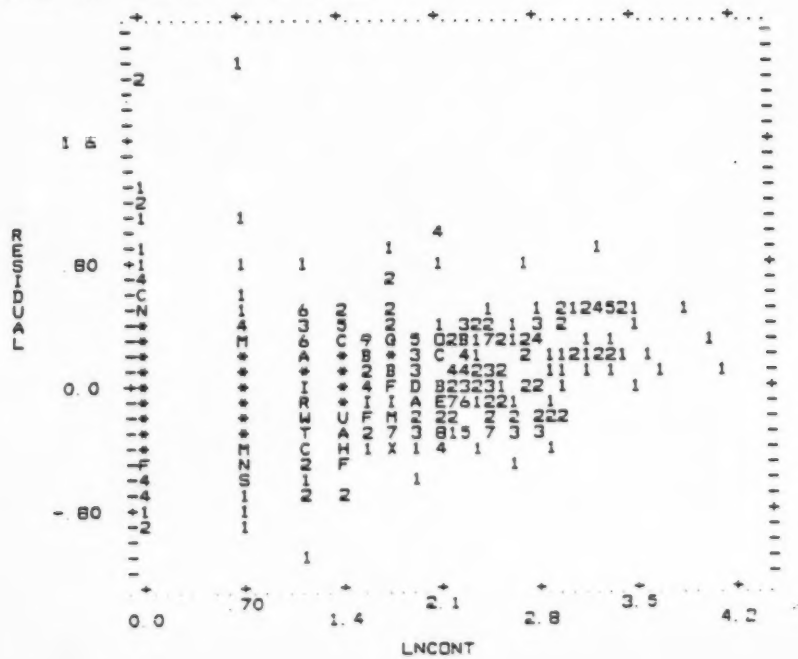


PAGE 12 BMDP1R FCL 20 CH RNEP

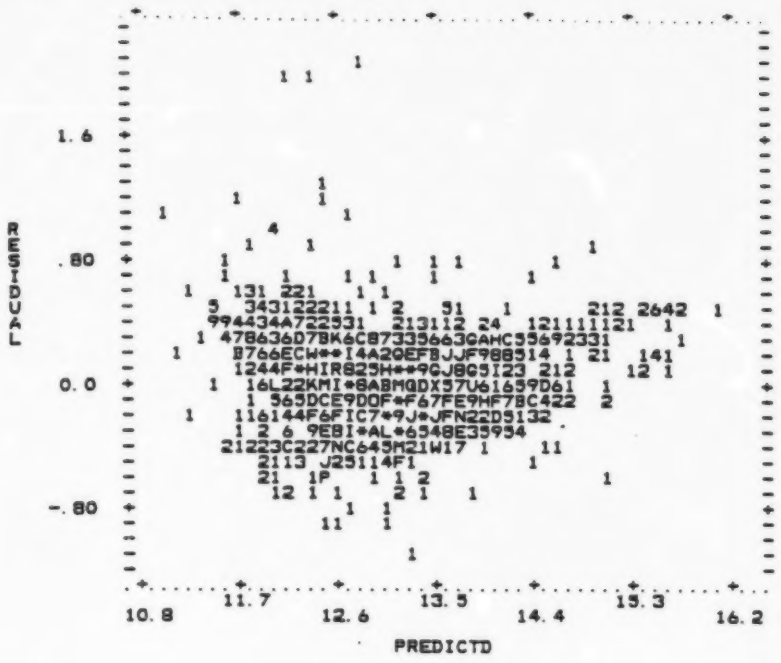




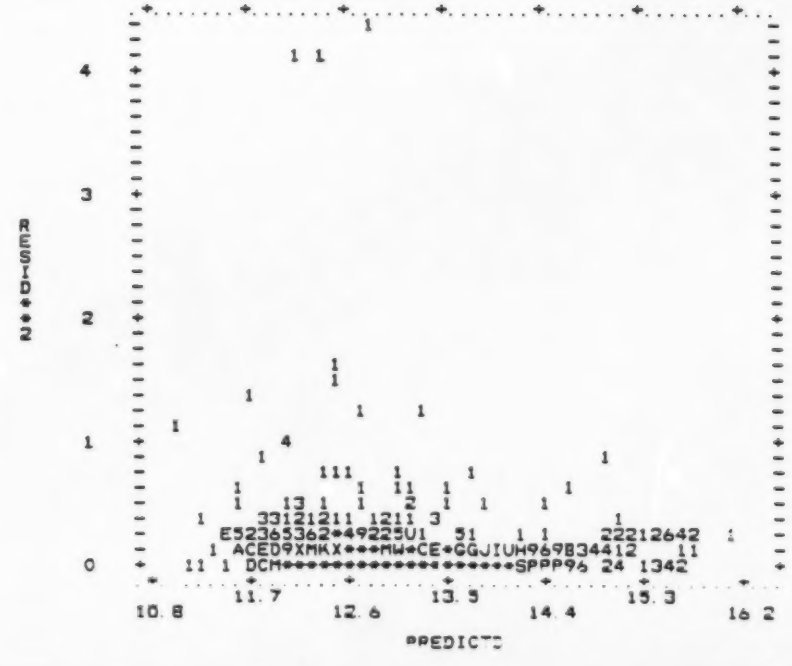
PAGE 10 BMDP1R FCL 20 CH RNEP



PAGE 7 BMDP1R FCL 20 CH RNEP



PAGE 8 BMDP1R FCL 20 CH RNEP



42.92 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 9879, CASE LABEL =

7.71 IS THE LARGEST STANDARDIZED RESIDUAL (IN ABSOLUTE VALUE) AMONG CASES WITH POSITIVE CASE WEIGHT. THIS OCCURRED FOR CASE NUMBER 5608, CASE LABEL =

.06 IS THE MAXIMUM VALUE OF COOK'S DISTANCE AMONG CASES WITH POSITIVE WEIGHT. THIS OCCURRED FOR CASE NUMBER 9879, CASE LABEL =
IF THIS CASE WERE OMITTED, THE REGRESSION COEFFICIENTS WOULD MOVE FROM THE VALUES REPORTED ABOVE TO THE EDGE OF A 2.05 PERCENT CONFIDENCE ELLIPSOID.

COMPARISON OF ESTIMATES OF REGRESSION COEFFICIENTS
(RELATIVE DIFFERENCE IS DIFFERENCE DIVIDED BY ORDINARY COEF.
STANDARD ERROR IS THAT OF ORDINARY COEFFICIENT.)

	ORDINARY LEAST SQUARES	OMITTING CASE WITH LARGEST COOK DISTANCE	RELATIVE DIFFERENCE	DIFFERENCE DIVIDED BY STANDARD ERROR
INTERCEPT	7.508542	7.477422	.0041	.4143
13 LNCONT	.861834	.862071	-.0003	-.0356
12 LNPOICO	.263217	.267811	-.0175	-.4729
11 LNKMDIS	.578016	.579430	-.0024	-.1595

NUMERICAL CONSISTENCY CHECK

RESIDUAL MEAN SQUARES ARE COMPUTED FROM BOTH COVARIANCE MATRIX AND RESIDUALS, AND RELATIVE DIFFERENCE (DIFFERENCE DIVIDED BY SMALLER OF TWO ESTIMATES) IS COMPUTED.

RESIDUAL MEAN SQUARES COMPUTED FROM		
COVARIANCE MATRIX	RESIDUALS	RELATIVE DIFFERENCE
.732973E-01	.732973E-01	.702813E-12

IN THE BIVARIATE PLOTS WHICH FOLLOW, A = 10 CASES, B = 11 CASES, ..., AND * = 36 OR MORE CASES.
AN * IN THE BORDER OF ANY PLOT INDICATES A MISSING VALUE OR A VALUE OUT OF RANGE.

FCL / CH / ANCP / 20'

PAGE 6 BMDP9R FCL 20 PIEDS CH RNCP

STATISTICS FOR 'BEST' SUBSET

SQUARED MULTIPLE CORRELATION	.88354
MULTIPLE CORRELATION	.93997
ADJUSTED SQUARED MULT. CORR.	.88342
RESIDUAL MEAN SQUARE	.073297
STANDARD ERROR OF EST.	.270735
F-STATISTIC	6906.61
NUMERATOR DEGREES OF FREEDOM	3
DENOMINATOR DEGREES OF FREEDOM	2731
SIGNIFICANCE (TAIL PROB.)	.0000

NO.	VARIABLE NAME	REGRESSION COEFFICIENT	STANDARD ERROR	STAND. COEF.	T-STAT.	2TAIL SIG.	TOLERANCE	CONTRIBUTION TO R-SQ
	INTERCEPT	7.50854	.0751067	9.470	99.97	.000		
13	LNCONT	.861834	.00665713	.861	129.46	.000	.963710	.71468
12	LNPOICO	.263217	.00971453	.179	27.10	.000	.972601	.03131
11	LNKMDIS	.578016	.00886661	.432	65.19	.000	.973137	.18122

THE CONTRIBUTION TO R-SQUARED FOR EACH VARIABLE IS THE AMOUNT BY WHICH R-SQUARED WOULD BE REDUCED IF THAT VARIABLE WERE REMOVED FROM THE REGRESSION EQUATION.

PAGE 7 BMDP9R FCL 20 PIEDS

CORRELATIONS OF THE ESTIMATES OF THE REGRESSION COEFFICIENTS

		LNCONT 13	LNPOICO 12	LNKMDIS 11
LNCONT	13	1.000		
LNPOICO	12	-.130	1.000	
LNKMDIS	11	.128	.085	1.000

SUMMARY STATISTICS FOR RESIDUALS

(CASES WITH POSITIVE WEIGHT)	
AVERAGE RESIDUAL	.0000
RESIDUAL MEAN SQUARE	.07329733
AVERAGE DELETED RESIDUAL	.0001
AVE. SQUARED DELETED RESIDUAL (PREDICTION MEAN SQUARE)	.07351836
SERIAL CORRELATION	.4989
DURBIN-WATSON STATISTIC	1.0017

12.22 IS THE MAXIMUM VALUE OF MAHALANOBIS DISTANCE AMONG CASES WITH [illegible] SECURED FOR CASE NUMBER 9879, CASE LABEL =

PAGE 4 BMDP9R FCL 20 PIEDS RNEP CH

NUMBER OF CASES READ. 11245
CASES WITH USE SET TO ZERO 8510
REMAINING NUMBER OF CASES 2735

SUMMARY STATISTICS FOR EACH VARIABLE

VARIABLE	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	SMALLEST VALUE	LARGEST VALUE	SMALLEST STANDARD SCORE
13 LNCONT	.85915	.79229	.922182	.00000	4.09434	-1.08
12 LNPOICO	4.73634	.54045	.114107	1.38629	7.68478	-6.20
11 LNKMDIS	6.19961	.59197	.095485	2.39790	7.18007	-6.42
14 LNDEP	13.07914	.79291	.060624	11.03972	16.57940	-2.57

VALUES FOR KURTOSIS GREATER THAN ZERO INDICATE DISTRIBUTIONS
WITH HEAVIER TAILS THAN THE NORMAL DISTRIBUTION.

PAGE 5 BMDP9R FCL 20 PIEDS

CORRELATIONS

		LNCONT 13	LNPOICO 12	LNKMDIS 11	LNDEP 14
LNCONT	13	1.000			
LNPOICO	12	.142	1.000		
LNKMDIS	11	-.141	-.103	1.000	
LNDEP	14	.826	.257	.292	1.000

ABSOLUTE VALUES OF CORRELATIONS IN SHADED FORM

13 LNCONT *
14 LNDEP **
11 LNKMDIS .-*
12 LNPOICO .-.*

THE ABSOLUTE VALUES OF
THE MATRIX ENTRIES HAVE BEEN PRINTED ABOVE IN SHADED FORM
ACCORDING TO THE FOLLOWING SCHEME

	LESS THAN OR EQUAL TO		.103
-	.103	TO AND INCLUDING	.207
+	.207	TO AND INCLUDING	.310
X	.310	TO AND INCLUDING	.413
N	.413	TO AND INCLUDING	.516
O	.516	TO AND INCLUDING	.620
*	.620	TO AND INCLUDING	.723
	GREATER THAN		.723